

Data Warehousing and the Web

Jim Davis, SAS Institute Inc., Cary, NC

ABSTRACT

A successful data warehouse is not solely dependent on surfacing data organized for decision support activities or controlling the processes by which the data warehouse is built. In fact, the success of the data warehouse is dependent on the ability of the end-user community to understand and gain access to the newly created information. The decision support user must be comfortable with the information that they are basing their decisions on.

The user, whether based in IT or the decision support departments, must be able to have answers to the following questions readily available: How do we gain access to the decision support data? Where does it reside in our enterprise? How and where was the data derived? When was the data loaded and when is it scheduled to be refreshed? What does all this information mean?

By using the SAS/Warehouse Administrator to define and build the data warehouse, the enterprise gains a new and significant asset – a central metadata store. The metadata repository contains all the process information and business rules associated with the movement of data from the operational environment to the decision support community.

This paper will examine methods for exploiting the information contained in the metadata repository via the web. Significant information can be surfaced for both the IT and end user communities.

INTRODUCTION

A number of factors are contributing to an increase in users who have access to strategic corporate information – most notably, data warehousing and web-based applications. Data is no longer restricted to a community of “power users”. Web-based applications are opening the door to new communities of users. Once thought of as independent corporate initiatives, data warehousing and web browsers have come together to form an effective approach to surfacing information to an expanding community of users. By surfacing decision support data within a web-based application, users are able to easily view and act on information that can move their organization forward and maintain the competitive advantage necessary in today’s market place.

Data Warehousing: What is it?

For years, organizations have been accumulating enormous amounts of data through transaction processing systems. Examples of these systems include order entry, airline reservation, and general ledger systems. While these systems play an integral part in maintaining the

smooth operation of day-to-day activities within an organization, they do little to provide the decision support community with the information they need to make effective decisions. Transaction processing systems are designed to “get information in”. They are very good at maintaining high individual transaction rates.

Decision support users are in the business of “getting data out”. Typically, this activity requires large amounts of summarized information in order to support an effective decision making process. Attempting this type of data access in the operational environment can result in disaster. Decreased performance and inaccurate results are just two of many potential pitfalls.

A data warehouse empowers the user community with the information they need to remain effective and competitive. Simply put, data warehousing is the process of turning data into information. It is the physical separation of the operational and decision support environments. Data is extracted from the operational environment and transformed into a new information store, or data warehouse. The information is organized with the usage and access patterns of the end user community in mind.

Exploiting the Data Warehouse

Once the warehouse has been built, the information is ready to be exploited by the user community. One approach is to provide the end user with a pre-built application that accesses the information in the warehouse in support of their decision support activity. While this application may be quite effective for a particular task, it does little to provide the user with a means to explore beyond the application’s defined scope. Competitive advantage can be achieved in a number of ways. One way is to provide our users with the means to search out and discover new ways of exploiting information.

So how do we “discover” new information in the warehouse? The ideal interface is one that is easily deployed to a wide range of users. Web-based applications are an ideal fit. They are low cost, easy to deploy, and provide a graphical user interface that appeals, and is available, to a broad range of users – the casual user to the power user.

As we’ll see, metadata represents the glue that binds data warehousing and the web.

THE ROLE OF METADATA

Organizations have had great success building warehouses with The SAS® System for a number of years. SAS has long been a provider of data access engines and a robust platform-independent transformation engine. The difficulty has come in trying to manage the

warehousing process across a diverse environment. The solution lies with metadata. The warehouse metadata repository is a facility that acts as a central storage point for maintaining the information necessary to define and control the warehousing process.

In his book entitled *Managing the Data Warehouse*, Bill Inmon states, "The heart of the architected environment is the data warehouse; at its nerve center is metadata. Without metadata, the data warehouse and its associated components in the architected environment are merely disjointed components working independently and with separate goals. In order to achieve harmony and unity across the different components of the architected environment, there must be a well-defined and disciplined approach to metadata."

SAS/Warehouse Administrator™ provides a framework for creating and managing the warehouse environment. The process is governed by a central metadata repository.

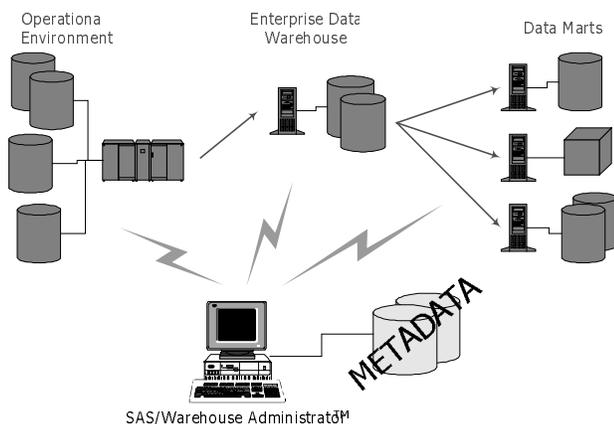


Figure 1 - Metadata, A Central Point of Control

Figure 1 shows the relationship between the warehousing process, SAS/Warehouse Administrator, and metadata. SAS/Warehouse Administrator acts as an agent in the enterprise to manage the flow of information from the operational systems into the decision support environment. Information on source data stores, necessary transformations, data movement, and the target warehouse is maintained in the central metadata repository. SAS/Warehouse Administrator provides a graphical front end to create and manage the metadata layer. Information contained in the metadata includes:

- Source data locations
- Definition of source data
- Access methods for source data
- Business descriptions of source data
- Business rules associated with data transformation
- Location and timing of transformation processes
- Warehouse table definitions
- Location of warehouse information
- Access methods for warehouse information
- Business descriptions for warehouse objects

- Owners and administrators of data and processes

BROWSING THE WAREHOUSE

The metadata repository represents a new and valuable corporate asset. Its potential extends beyond that of a data dictionary, case tool, or modeling tool. It contains include business rules and definitions associated with the warehouse, as well as the operational systems used to derive the warehouse. The value of metadata can be extended beyond the community of users responsible for the creation and maintenance of the warehouse.

The metadata repository can be used as a road map for navigating the enterprise. Web-based applications can read the metadata, use it to navigate the operational and decision support environments, identify objects, and because the metadata contains all the necessary connection information, surface data to the end-user tool of choice.

Warehouse Viewer™

Warehouse Viewer is a sample application that is available with SAS/IntrNet™ software that allows users to browse the contents of a warehouse. With Warehouse Viewer, users can access the contents of a warehouse that is being managed by SAS/Warehouse Administrator. End users can use their web browsers to locate tables, charts, graphs, and documentation associated with the warehouse. Users are provided both technical and business metadata.



Figure 2 - Warehouse Viewer, Directory View

Figure 2 shows how the Warehouse Viewer depicts a directory view of the "Corporate" warehouse. In this scenario, the user is accessing a SAS IntrNet server which has access to exported data from SAS/Warehouse Administrator-maintained metadata. In this view, the user is looking at the subject areas available. As the cursor moves across warehouse objects, additional information on that object is displayed in the boxes in the lower part of the screen.

Physical City	Physical State	Physical Country	Physical Location	Customer Name	Last Order At
Southwest	ARIZONA	US	LAJAVE	SMITH-AND INC	000179-03-00-00
Northwest	IDAHO	US	LAJAVE	SMITH-AND INC	000179-03-00-00
West	OREGON	US	LAJAVE	SMITH-AND INC	000179-03-00-00
West	WASHINGTON	US	LAJAVE	SMITH-AND INC	000179-03-00-00
West	BRITISH COLUMBIA	CA	LAJAVE	SMITH-AND INC	000179-03-00-00

Figure 3 - Warehouse Viewer, Surfacing Data

Once the user has identified their area of interest, they can then request access to that information. *Figure 3* shows data from the customer subject area of the warehouse. In this case, the SAS/IntrNet server uses the metadata to obtain the information necessary to gain access to the customer detail table. The server then surfaces the data 50 records at a time to the end user's browser. At this point, the user may continue reviewing the data via the browser environment or may request that the data be downloaded from the warehouse for further manipulation using their desktop tool of choice.

A search facility is also included. Users may search the warehouse by keyword or they may choose to search for objects by owner or administrator.

Technical users also benefit from this application. The Warehouse Viewer allows administrators to focus on metadata that defines the warehouse. For example, warehouse structure and physical file locations.

JWarehouse Viewer: Java-Based Warehouse Viewer

(Note: The following application was under development at the time this paper was being written. The application name is subject to change.)

JWarehouse Viewer provides a similar feature set to that found in the CGI-based Warehouse Viewer. JWarehouse Viewer is a Java applet. No software, other than a Java-enabled browser must be installed on the user's platform. Again, focusing on the needs of a browser-based user community, this application provides an end user with exploratory access to a warehouse by exploiting the metadata maintained by SAS/Warehouse Administrator. While the warehouse viewer provides information for both the technical and business communities, this Java application focuses in on the needs of the business user, not the administrator of the warehouse.

Functionality includes:

- Browse the warehouse with objects organized by subject, type, or owner.
- Locate objects on the appropriate server.
- Search the warehouse for objects meeting user specified criteria.
- Display objects using a default model/view.

- Display metadata associated with the object.
- Transform the object (via filter, procedure, etc.).

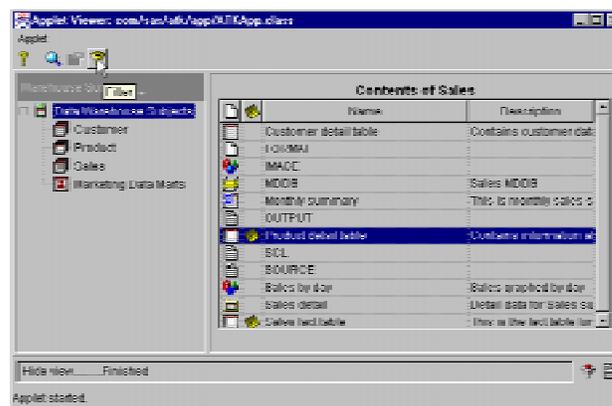


Figure 4 – JWarehouse Viewer, Main Screen

Figure 4 illustrates how JWarehouse Viewer depicts the warehouse to the user. The user is presented with a splitter panel that segments the window into two areas. The left hand side of the screen shows the warehouse and lists the subject areas within that warehouse. In this example, the Sales subject area was chosen and the right hand side of the screen lists the contents of that subject area.

The tool bar is segmented in two areas – tools that operate on the warehouse and tools that are associated with the selected object. All tools have textual tool tips associated with them. In *figure 4*, the cursor is placed on the filter icon and the associated tool tip is displayed.

There are three components located at the bottom of the main window that provide the user with status information. The components are:

1. *Status Line*. This provides textual information about the commands currently executing.
2. *SAS Pyramid*. If the pyramid is spinning, there is at least one command executing.
3. *Status Window Button*. When selected, a window containing an activity log will appear.

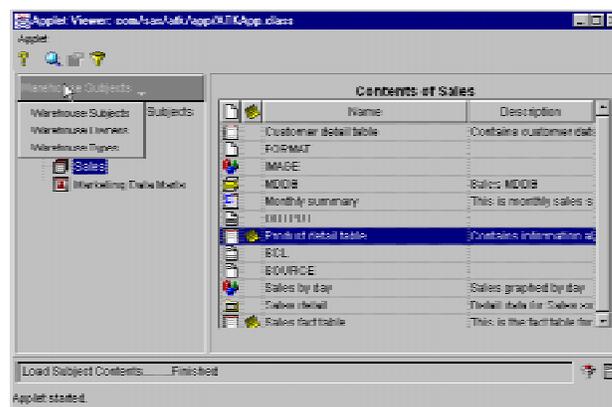


Figure 5 - JWarehouse Viewer, View Selector

Figure 5 illustrates how the user can change the way in which warehouse information is organized. There are three possible settings:

1. *Warehouse Subjects*. The tree component is subject oriented. Items in the tree are the unique subjects registered in the warehouse.
2. *Warehouse Owners*. The tree component is owner oriented. Items in the tree are the unique owners listed in the warehouse. Further expansion reveals objects owned by a particular individual.
3. *Warehouse Types*. The tree component is type oriented. Items in the tree are unique types registered in the warehouse. For example, if "GRSEG" were selected, the view would contain all objects in the warehouse of type "GRSEG".

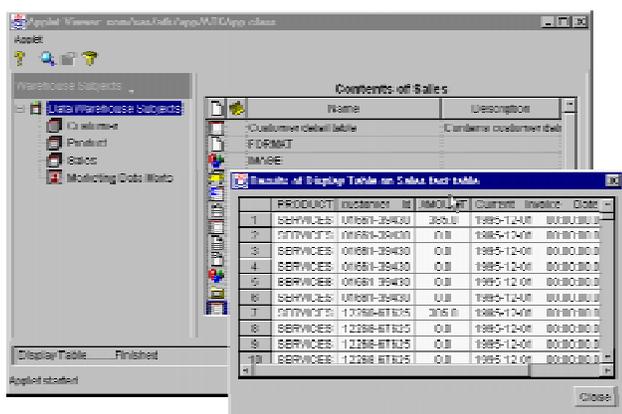


Figure 6 - JWarehouse Viewer, Viewing an Object

An object can be selected and viewed. Each object has an associated default view. Figure 6 shows the default representation of a standard data table object. The viewer launches windows external to its main window. This allows the user to view multiple objects at a time.

It is important to note that JWarehouse Viewer provides a multi-threaded interface. The viewer creates a separate thread to handle each request from the user. The user is afforded the luxury of initiating multiple tasks, some of which may require significant execution time, and may continue working without a lock being placed on the user interface.

CONCLUSION: WHAT'S NEXT?

This paper has concentrated on technology available from SAS Institute today. There are developments on the horizon that may very quickly enable the end user community to reap additional benefit from web-based access to the data warehouse.

Version 7: Metadata Integration

In the previous examples, browsing the warehouse is dependent on metadata that is created and maintained by SAS/Warehouse Administrator. Version 7 ushers in technology that brings together metadata from a number of diverse sources. The metadata repository in version 7 provides a common metadata facility for SAS applications.

This means that metadata you maintain in support of one product is available to other SAS products without conversion or import/export. In addition, facilities will be in place to allow for metadata exchange with other non-SAS repositories that comply with current metadata interchange standards.

Web-based applications will be made that much easier, being able to derive their information from a single point in the organization.

REFERENCES

Inmon, W.H., Welch, J.D., Glassey, K.L., (1997), *Managing the Data Warehouse*, John Wiley & Sons, Inc.

ACKNOWLEDGMENTS

Thanks to the following members of SAS Institute Inc., Cary, NC for their help in preparing this paper:

Barbara Walters
Catherine Corbett
Bryan Boone

AUTHOR CONTACT

Jim Davis
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
(919) 677-8000
sascjd@wnt.sas.com

The SAS® System is an integrated system of software providing complete control over data access, management, analysis, and presentation.

SAS/Warehouse Administrator™, SAS/IntrNet™, and Warehouse Viewer™ software are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

The Institute is a private company devoted to the support and further development of its software and related services.

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