Bringing Business Information to AT&T Network Systems Through a Data Warehouse

Stacey J. Gelman
W. Douglas Peck

The concept of *data warehousing* originated from the observation that the systems used to run businesses on a daily basis differ fundamentally from those employed to help plan and develop future businesses. For example, operational systems are generally focused on specific functional views based on the needs of a single aspect of the business. However, managers need information that shows relationships, trends, and correlations about different kinds of data, integrating several functions into a broader view. Historically, systems and manual processes were established to gather management data from the various operational data sources—one for each kind of decision. Extracting and combining such data from different systems is time consuming and often leads to inconsistent results. Users must accommodate printed reports, manual reentry of data into spreadsheets, and significant rework to produce summary reports that match the way they manage the business. Furthermore, by the time some of these reports are ready, the data are no longer current. The *Warehouse of Information for Network Systems (WINS)* provides needed information to AT&T Network Systems (AT&T-NS) managers world wide. WINS transforms operational and financial data into consolidated business views that are used to analyze certain activities and to make management decisions. This paper discusses the importance of WINS to the business management strategy of AT&T-NS, the WINS technical architecture, the status of WINS, and plans for future implementation.

Introduction

Data warehousing is the result of the simple but significant observation that the operational systems used to run AT&T Network Systems (AT&T-NS) businesses on a daily basis are fundamentally different from those that help plan and develop AT&T-NS future businesses.[1, 2] Operational systems are generally focused on a specific functional view (for example, order entry, billing, and general ledger information) based on the needs of a single aspect of the business. However, managers need information that integrates various functions into a broader business view that shows relationships, trends, and correlations about different kinds of data.

Historically, to acquire the necessary management data, systems and manual processes were established to gather it from the various operational data sources—one for each kind of management decision. Extracting and combining data from different systems, however, is time consuming and often leads to inconsistent results. Users are faced with printed reports, the necessity for manual reentry of data into spreadsheets, and significant rework to produce summary reports that match the way they manage the business. By the time some of these reports are ready, the data are no longer current.

The existing business process environment needs to reflect the following four new requirements to ensure continued high performance:

- Quick reaction to the changing external customer world,
- Provision of only needed information in a timely manner with absolute data integrity,
- Focusing of resources from many areas of the business to solve problems efficiently, and
- Integration of the operational and financial views of the business.

The *Warehouse of Information for Network Systems (WINS)* provides needed information to AT&T-NS managers world wide. WINS transforms a vast collection of operational and financial data into consolidated business views that are used to analyze activities and make management decisions.

This paper discusses the importance of WINS to the business management strategy of AT&T-NS. It further discusses the WINS technical architecture, the status of WINS, and plans for future implementation.

Panel 1. Abbreviations, Acronyms, and Terms

AMP—access module processor
AP—application processor
AT&T-NS—the Network Systems business unit of AT&T
MIPS—millions of instructions per second
MPP—massively parallel processor
ODBC—Open DataBase Connectivity, a programming interface for Windows* applications to access databases on a network
PE—parsing engine
TCP/IP—transmission control protocol/Internet protocol, a communications protocol used to internetwork dissimilar systems
WINS—Warehouse of Information for Network Systems

Why a Data Warehouse?

To administer a large business, managers must have information that shows the broad status of its operations, key cross-operational factors, and important correlations among the factors. Managers use this information both to redirect resources strategically and to make process changes to improve business operations. Consider, for example, how a business might examine its delivery performance—that is, the ability of the business to deliver customer orders on time. The operational system for order entry carries the status of each individual order. However, the degree of customer satisfaction can be judged only by grouping all the orders together and examining broad trends. To perform root cause analysis on *why* deliveries are not meeting expectations, the order data must be correlated against information describing the state of the various providers in the order stream.

In a highly computerized operational environment, virtually all the data required for these decisions are captured in operational databases supporting specific functional areas. The good news is that such databases can be readily accessed. The bad news, discussed next, is that access is both technically difficult and fraught with some interesting business problems.

Before the advent of the data warehouse, people attempted to address the need for information to support management decision-making on a per-decision basis. Individual decision support systems and processes were created that gathered the appropriate data from the operational databases, consolidated them, and then formatted reports to support the given decision process. Many well-known applications can perform these functions, as can the hundreds of semi-manual processes that people have developed for their own areas. These applications and processes are valuable, but they are characterized by at least the following four separate problems:

- *High cost.* Implementing unique systems and processes for each decision area requires significant development resources. Additionally, in many cases, the resulting processes are only semi-automated (involving re-keying data into spreadsheets and manual correlation of data from different sources), so the ongoing operational cost of obtaining the data is quite high.
- *Delayed results.* People must be placed into the semi-automated processes. Thus, a significant delay results between when the actual events that create the data in the operational database occur and when they are reflected in the decision support data. As a result,

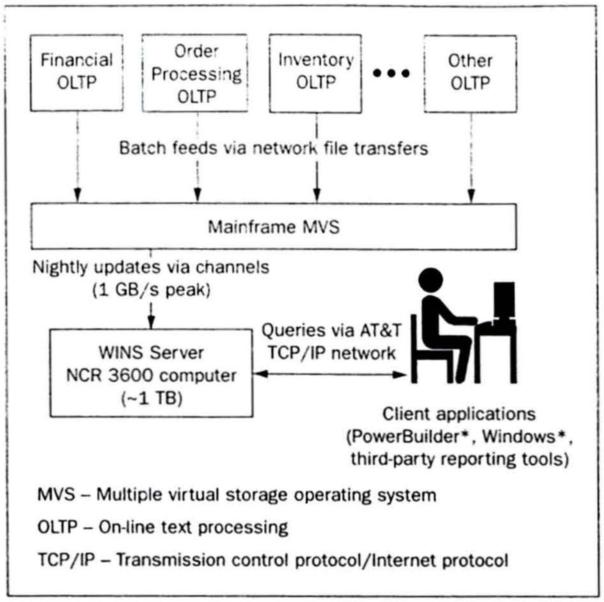

Figure 1. WINS uses a two-tier client-server architecture. The user interface and client applications run on the user's PC, which is connected through a network to a very large NCR 3600 database server. Data, extracted on a regular schedule from multiple operational systems, are sent in bulk over the AT&T internal network to a mainframe computer. Following the loading of new feeds, application processes are run on the server to preprocess and summarize the information, updating the tables that will be accessed by the client applications on the user's PC.

managers do not have the up-to-date data they need to react quickly.

- *Uncoordinated data for different decision processes.* Data for different classes of decisions are separately compiled. Such compilations can be obtained from different operational sources. In addition, they may reflect operational conditions from different periods of time. For example, an order status summary may show delivery fall-downs due to an out-of-stock condition, while an inventory summary—taken a week earlier or later—may not reflect the same inventory problem.
- *Inflexibility.* Often, the most important questions are the ad hoc management questions that are suggested by examining the original data. If a long cycle is necessary to develop a new process to answer an ad hoc question, the answer may no longer be useful. Even worse, the need to obtain an answer quickly may lead to assumptions about the data, which turn out to be incorrect.

To move beyond this world of multiple independent and semi-automated decision systems, AT&T-NS is building WINS, a unified repository of data drawn from the various operational systems that run the business. WINS data are kept current and complete and offer a powerful environment that allows the rapid creation of complex applications and even ad hoc queries against the data. Thus, WINS provides AT&T-NS managers the information they require to correctly make the decisions needed to run the business efficiently.

WINS Basic Principles. The WINS data warehouse concept consists of the following three processes:

- *Gathering*, which seeks and collects data from all the operational systems running the business;
- *Consolidation*, which reconciles and stores the data as a unified structure based on an integrated data model; and
- *Reporting*, which usually consists of either prebuilt applications (generally created using high-power tools like PowerBuilder* or Visual Basic*) or ad hoc queries (generally created using tools like Business Objects*).

A very reasonable, frequently asked question is, Why does a physical data warehouse exist when the data are already available in operational systems? It is not immediately clear why this duplication of data is necessary. Another way of warehousing data is to establish only a logical data warehouse, which is essentially a directory indicating in which operational database the desired information is located. Collection and consoli-

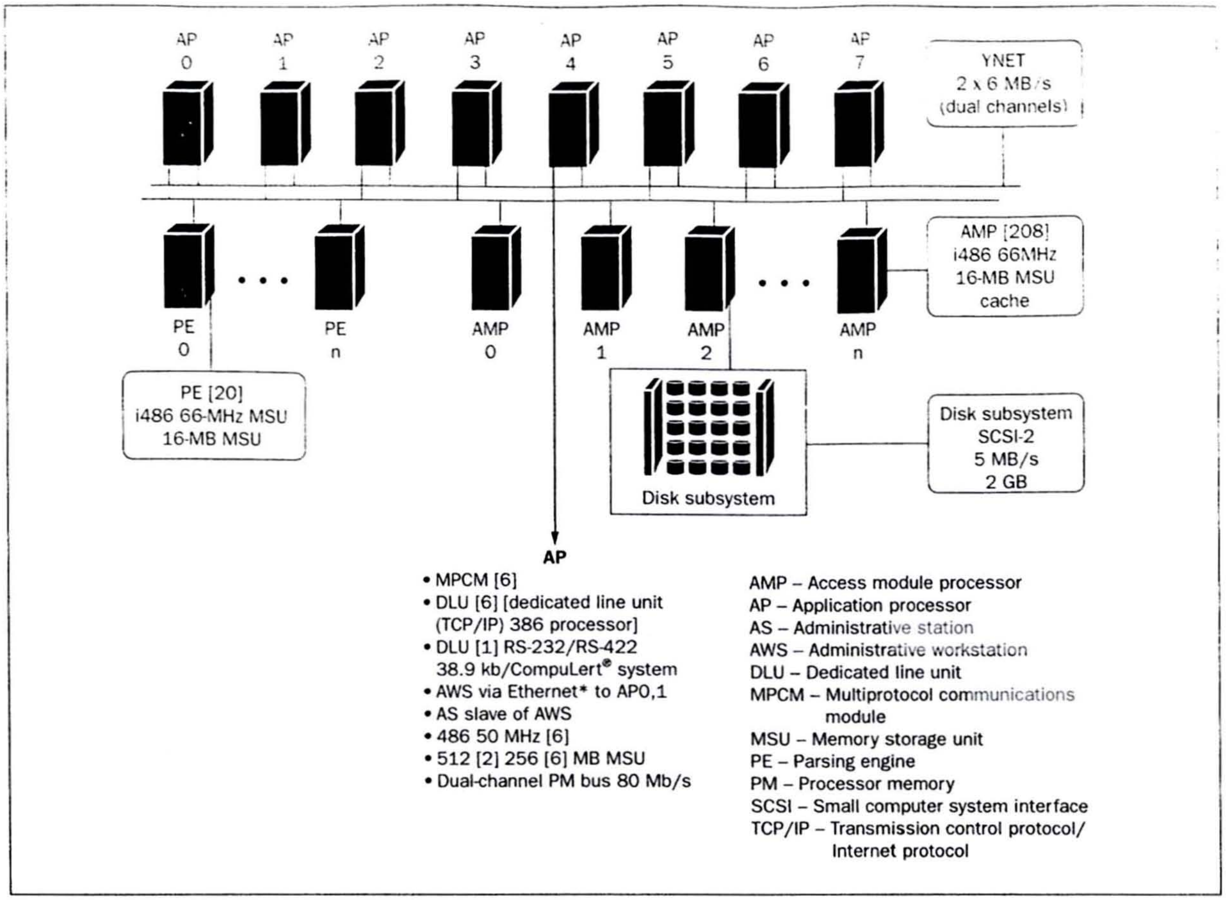

Figure 2. The NCR 3600 database engine consists of several hardware processors linked through a high-speed intelligent interconnect called the *Ynet*. A request from a PC application is routed to a parsing engine (PE), which decodes the request and constructs an optimized plan for executing it. The plan consists of a series of instructions for each database processor (AMP). The AMP executes instructions sent to it from the PE, returning its part of the processed data over the Ynet. The data are then returned over the network to the user's PC for display. The NCR 3600 computer enables queries to be performed on very large amounts of data that would otherwise run too slowly to be practical.

dation of the data from these sources on demand would then serve the same function as the data warehouse. This is not a reasonable strategy, however, based on the following two factors:

- Realtime consolidation is a complex task, and
- Complex queries from the logical data warehouse could lock up an operational database for time periods that would seriously interfere with the operational applications working off the same database (for instance, a 30-second lockup on an order database when a half-second response is required from order entry systems would be disastrous).

The availability and ready access to data in the warehouse brings with it requests for data use that should still be served by an operational or local system. The data, by definition, are not completely current. For example, the status of an order for a particular customer should be obtained by accessing the operational system in which up-to-the-minute data are available.

To assess the applicability of a business need to WINS, a set of operating principles was developed. Each request for new data types and/or associated reporting (either prebuilt applications or ad hoc queries) is evaluated with respect to the following five principles to arrive at either a *go* or *no-go* decision:

1. *The warehouse is read-only.* WINS applications (including reporting and ad hoc queries) provide only information. No direct data entering or updating of operational business data into WINS is carried out.
2. *The warehouse is not the sole or official source of the data it contains.* Business data in WINS are created and maintained via periodic feeds from operational databases.

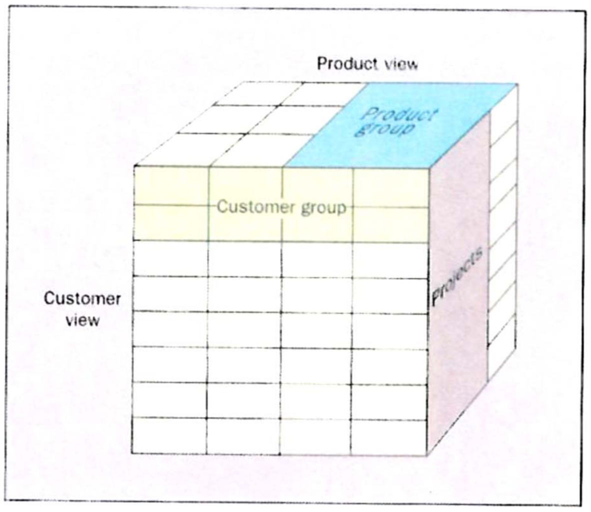

Figure 3. An important capability provided by the data warehouse is the ability to view data along many different dimensions or views. For example, it may be desirable to view details on revenue by sales organization, product, customer, project, business unit, country, and so on. The data warehouse architect must pay particular attention to designing summary structures that can support these multidimensional views. Furthermore, these views may change from time to time (for example, when the business reorganizes), and the ability to restate massive amounts of data quickly into the new views is important.

3. *The warehouse contains data for decision support, not for routine operations.* The frequency of updates to WINS data is established for each type of data as the longest period that:
   - Sustains the decision support application data needs; and
   - Ensures no key data are missed (for example, no important operational data have come and gone in between updates).
4. *The warehouse data are consistent between different reporting views (or applications) of that data.* All views of the same data from the same source will always be kept synchronized.
5. *The warehouse is used as a repository for consolidated views throughout AT&T-NS.* Priority is given to applications requiring data consolidation (reconciled data from multiple sources) and that serve AT&T-NS.

Assessing requests for new applications and data against these principles ensures that the needs of the user community are best served. A handful of warehouse requests have been rejected because either the requests required information to be updated (the entry of information into the warehouse violates principle 1) or they required realtime access to information (in violation of principle 3). In both cases, the operational system was used to fulfill the business need.

WINS Architecture

The next seven subsections discuss and examine the specifics of the WINS architecture and how it transforms data into consolidated business views.

Hardware/Software Platform. WINS is implemented using a two-tier client-server architecture as shown in Figure 1. The user interface and the client applications run on the user's PC, which is connected through a network to a very large database server, the AT&T Global Information Solutions (now NCR) 3600 computer. Data are extracted on a regular schedule from multiple operational systems daily, weekly, or monthly. The data are then sent in bulk over the AT&T internal network to a mainframe computer. The mainframe is used as a convenient staging area for receiving the feeds for nightly bulk loading into the server. Following the loading of new feeds, application processes are run on the server to preprocess and summarize the information, updating the tables that will be accessed by the client applications on the user's PC.

When a user requests information, the PC client application prepares one or more information requests from the database server and transmits the request(s) to the NCR 3600 server over the AT&T network. The computer performs the data extraction and manipulation processes and returns the data over the network to the user's PC, where they are formatted and displayed.

Client applications are written using the PowerBuilder* tool set, and they run on the user's PC under Microsoft* Windows*. Vendor software, such as Business Objects*, can also be run on a PC to access the data warehouse. However, such software must be packaged and controlled as part of an AT&T application so that security and performance can be managed, even though a standard ODBC driver is used to make client requests to the server. A variety of TCP/IP communications packages is used to connect the PCs to local area networks and the AT&T wide area network.

The NCR 3600 server is configured with eight *application processors (APs)*, each of which is a general purpose computer running the UNIX* system. These computers can be used to host applications that implement shared business functions, thus further shielding users from direct access to warehouse data. Most of the warehouse processing, however, takes place on the specialized database hardware of the NCR 3600 computer.

The NCR 3600 database engine consists of many hardware processors linked through a high-speed intelligent interconnect, the *Ynet*, as shown in Figure 2. A request from a PC application over the network is received by the server and routed to a *parsing engine (PE)*, which decodes the request and constructs an optimized plan for executing it. The plan consists of a series of instructions for each database processor called an *access module processor (AMP)*. Each AMP has disks directly attached to it. An AMP executes instructions sent to it from the PE, returning the AMP's part of the processed data over the Ynet. (Data may also be passed from one AMP to another depending on the database operations to be performed.) The data are then returned over the network to the user's PC for display.

The architecture of the NCR 3600 computer makes it possible for all the database processors to work in parallel to complete a user's request, which is why this type of machine is called a *massively parallel processor (MPP)*. The NCR 3600 computer enables queries to be performed on very large amounts of data. The queries would run too slowly on conventional processors to be practical.

The NCR 3600 server residing in WINS is currently configured with 20 PEs, 208 AMPs, and 832 gigabytes (GB) of disk space controlled by the AMPs. Another 416 GB is currently being added to support more application data. Given the pattern of data access anticipated, the ratios of these elements are determined based on NCR heuristics, which have been effective. The combined raw processing power of the hundreds of CPUs on the WINS NCR 3600 computer (MIPS) exceeds the total power of AT&T's installed base of conventional mainframe computers. If all the bits of information that can be stored in the WINS database were printed, the stack of paper would stand about 100 miles high. (Other configurations of the NCR 3600 computer currently manage up to 4 terabytes (TB) of information in production. The architecture of the next generation of the machine, called the NCR 5100M, eliminates the specialized hardware processors and is designed to be expandable to 100 TB.) Despite the computer's large size, the predicted interval between failures that would result in loss of data is 59 years, and a target of 99.5-percent availability has been exceeded.

The need for initial data capacity of more than 200 GB for financial data was estimated based on the size of feeds to the general ledger, plus necessary fallback and system overhead. In early 1994, when the database was selected and configured, AT&T-GIS (now NCR) was the only vendor willing to support the data volume of AT&T-NS using the NCR 3600 computer. Since then, both ORACLE Corp. and Informix Software Inc. have increased their advertised capacity to approximately 500 GB.

Key Issues in the WINS Architecture. The following issues were critical in determining the architecture of WINS:

- Views of the data to be provided,
- Amount of data,
- Needs and abilities of the user population,
- Integrity requirements on the data, and
- Security to be provided.

While these issues are crucial to most computer systems implementations, the choice of their architectural solutions is particularly important for a data warehouse. In addition, such solutions may differ from the approaches needed for operational systems.

Views of the Data. Operational systems usually enable access to the information they store by their primary users. An order entry system, for example, may be designed for access by order number or customer name. An important capability provided by the data warehouse is the ability to view data along many different dimensions or views (see Figure 3). For example, the order information may need to be viewed according to what currency will eventually be used for payment, a question that may not be of concern to users who enter and track orders in the operational order management system. It may be desirable to view details on revenue by sales organization, product, customer, project, business unit, country, and so on. The data warehouse architect must pay particular attention to designing summary structures that can support these multidimensional views. Furthermore, the views may change from time to time (for example, when the business reorganizes), and the ability to restate massive amounts of data quickly into the new views is important.

To provide multidimensional access to the data, the first and most significant WINS architectural choice was the use of an MPP platform, which allows all the data to be stored on a single computer. An alternative was either to partition the data across multiple platforms—putting, say, all order data on one computer, all financial data on another, and so on—or to place all data for each geographic region together on one server. This alternative can provide good performance for queries that ask to view the data along the chosen partitioning. It may result in poor performance on all other views, however, due to the intermachine processing that must be carried out over much slower interfaces than those provided in the MPP arrangement. Because much of the business value of the data warehouse comes from its ability to provide multidimensional views, the MPP approach was preferred.

The second architectural decision is to create presummarized views of the data to improve query response. These view tables consist of records that "roll up" or summarize all the detailed data that match each unique combination of code values across the various dimensions. That is, all the revenue for each unique combination of organization, product, customer, and country values may be summarized into one record. This dramatically reduces the number of records that must be accessed to satisfy a user query.

In addition, the need to correlate information from different parts of the business—consider, for instance, the relationship between customer satisfaction and cost-of-quality—means that the summary structures for different types of information must be accessible along the same set of dimensions so that they can be joined. To provide for views of the data along consistent dimensions, WINS uses a set of shared tables that define common corporate data items, such as financial accounts and customers. These tables are centrally updated and also used by the operational systems that send the data to WINS.

Amount of Data. A data warehouse, which can be accessed on line, has a very large capacity. Typically, hundreds of GB and even up to a few TB of data are available (the size of the maximum available configuration is expected to grow rapidly), and all or nearly all the data are available to users through queries. Thus, performance engineering is absolutely critical, even though tremendous computing power may be available to support the query process. Performance analysis must be done on the processes that load, presummarize, and back up the database tables and also on the processing done for user queries. If ad hoc queries are permitted, estimates of the query load will be inexact at best, and the load placed on the computer by users may vary dramatically.

The primary WINS architectural challenge is how to deliver information within a satisfactory timeframe. Acceptable response may vary from seconds to hours, depending on a user's expectation of the amount of work being requested. Most WINS queries are completed in as quickly as a few seconds and up to a few minutes. Response time is optimized by choosing database summary structures that can be used to satisfy most user queries quickly without requiring massive detail processing. These structures presummarize the data along the most requested viewing dimensions using batch processes that are run whenever new data feeds are loaded into WINS.

Multiple summaries of the same underlying detail information may be needed to support different types of queries. "Drill-down" into the detail behind the summarized results can then be provided efficiently. The monitoring and logging of user activity can be employed to analyze which queries are run frequently so that additional summary tables can be created to support the queries more efficiently. However, the processing needed to create the summaries may be extensive (up to six hours), so a balance must be maintained between processing time spent presummarizing and time spent responding to queries.

Similarly, control over user queries—particularly over ad hoc queries—must be maintained to prevent a single user from overloading the WINS computer, which could happen if an inefficiently structured query is launched against large database tables. Users generally access only views of the data that have been optimized for good response. Sophisticated users, however, might need direct access to the detail data and they may initiate a query that runs very slowly. Queries can be monitored for the time and intermediate storage they require. Any difficult and highly complex queries can be rescheduled for overnight processing if no more efficient access to the information can be provided.

In many cases, WINS users—employing their own tools—simply run queries to download selected data to spreadsheets or local databases for further analysis. On-line analytical processing tools provide another significant opportunity to offload this type of analysis onto local user platforms.

Feeder loads and presummarization processes are scheduled and run regularly to minimize peaks in the processing load. Other maintenance activities are also carefully scheduled. The WINS database on the NCR 3600 computer is currently backed up to tape even though the feeders used to create it are also saved and could be used to re-create the database in an emergency. Executing such a procedure, however, would take days. As the amount of data on the computer grows, full back-ups may eventually become impractical due to the time required to complete them. AT&T-NS is already moving from full weekly back-ups to full quarterly back-ups. Scheduling of batch processing activities, which slow down the computer with updates, is complicated by the need for global access and a 24-hour business day.

User Needs and Abilities. Unlike operational systems, which are typically designed to meet the fairly uniform needs of a limited community of users, WINS is used by people in different parts of the business. Individuals vary widely in both their information needs and sophistication as computer users. Some users want regular reporting of standardized high-level summary information. Others need to construct intricate queries frequently using large amounts of detail data to address specific questions. A variety of user interfaces and tools is required to address the needs of these different sets of users even though the users may be accessing the same underlying data.

To meet the disparate needs of the various WINS users, multiple interfaces are provided. These interfaces include the following two different types: general-purpose database access tools, such as Business Objects*, that can provide complex ad hoc data access and report design capability for more sophisticated users; and very simple AT&T-built screens that allow users to request predefined canned queries for routine review. WINS is designed to have a single logical data model with all users sharing the same underlying data. The data may be presented in a variety of ways, however, by means of multiple interfaces tailored to each set of user needs.

Data Integrity. WINS contains information from many different source systems. The sources may be inconsistent, both because of business processes—such as the need to process orders even though some of the information on the order may be incomplete or incorrect—and because the information extracted from the warehouse is at a lower level of detail than has previously been provided outside the source system.

For example, information flowing between many AT&T systems is carefully validated through an extensive set of controlled tables to ensure that the information is correct. The information extracted for WINS, however, has a lower level of detail than the source system normally sends to other systems for operational processing. Thus, this information is not already being validated with the same rigor. WINS makes visible the inconsistencies between the data being used in different areas of the enterprise, in some cases confirming problems that have been suspected but never quantified.

The timing of data extraction from the source systems, which can affect consistency, is also an issue. Additionally, the source systems may use the same name but a different definition for some business information. For example, does "ship date" mean shipment from the factory, the assembly center, or a distribution warehouse? Does it mean the customer's requested, committed, or actual ship date? The answers may be different for the data sent to WINS from different source systems. In some cases, these differences in definition are historical, but in other cases they reflect genuine differences in user needs. Resolving the business data definitions is a difficult and time-consuming process.

WINS accommodates inconsistencies in the data sent to it from the various source systems. In general, WINS does not attempt to correct the data to maintain consistency with the source system. Ultimately, improvements in data quality must be made in the business processes and systems that create the data. WINS does attempt to correlate the data and flag inconsistencies, however, and the WINS data model actually reflects some of the inconsistencies that are expected to occur. (An important decision in creating the data model was to represent the data and their relationships currently available within the source systems rather than as an idealized version of what data should be available.) The use of a shared logical data model, as well as the common processing of data from source systems into tables for use by all the viewing applications, help ensure that the user view of the data is as consistent as possible.

Security. Design for security is important for every system. For a data warehouse, the concentration of proprietary business information makes security particularly important. The wide access to the data warehouse within the company also raises new business policy questions about whom should be allowed to see what data. Coupled with the need for multidimensional access to the information, wide access to the data warehouse can make design of the summary structures even more complex because business policy may authorize access to data at a summarized but not a detailed level or vice versa. WINS security includes controls on access to the platform (through an authentication server), use of permissions on applications and screens within applications, and control of which underlying data can be accessed. Various controls are implemented to log and monitor access and activity.

Back-up and recovery strategies for business continuity and disaster recovery may be significantly different for a data warehouse than for operational systems. This is due to the sheer size of the database and the availability of back-up hardware should the primary data warehouse not be available. The WINS computer hosts only decision support applications not critical to the day-to-day operation of the business, and it contains only data obtained from other systems. No data are created on the platform. Mirroring is used to avoid loss of data due to disk failures, and maintenance is scheduled to minimize the interval before defective components are replaced.

Status and Plans

The WINS concept was developed early in 1994 in conjunction with AT&T-NS financial users who were seeking better reporting mechanisms of consolidated financial and operational views for the division's business managers. The first application went live in October 1994, focusing on providing visibility into inventory and enabling the following capabilities:

- Determining material availability and the number of days of supply,
- Determining overstock and understock conditions,
- Comparing inventory data with operational financial sub-ledger and ledger systems, and
- Providing trend data for the production planning process.

The first live application of WINS required the consolidation of information from inventory, factory, and financial systems, as well as marshaling centers.

The overall approach to WINS is to phase in the addition of types of data and associated applications over a three-year period. In 1995, more details were added to data already coming into the warehouse and to the following types of data:

- Order,
- Billing,
- Sales,
- Invoice,
- Accounts receivable,
- Revenue,
- Sales forecast,
- Expense,
- Customer, and
- Product.

These data types require over 58 feeders—that is, the sources of the information are spread over 58 operational systems. For example, marshaling data, used in the inventory application, must be consolidated from seven different operational systems. Redundancy exists in the data needs of the different application views into the warehouse. For instance, accounts receivable data are used by four applications currently on WINS; order information is used by three applications.

The following list depicts the principal highlights of the applications providing data access and reporting to users:

- *Business performance.*
  - Sales tracking for weekly and monthly results for both direct and indirect sales. Direct sales information includes sales and the cost of sales data for basic gross profit analysis. To support sales tracking, sales history data have been maintained on WINS since 1991.
  - Demand tracking and reporting to project future revenue based on customer-requested ship dates. To support this analysis, WINS maintains a rolling 13-month cycle and a 36-month "snapshot" of orders having data categories, which include unfilled, canceled, and rescheduled orders, as well as deletions.
- *Order visibility.* Performance analysis for order status, metrics, cycle times, and forecasts.
- *Financial visibility.* Reports expenses at summary levels and enables users to scrutinize the details, providing the needed visibility to facilitate expense management.
- *Forecast analysis.* Trend analysis based on a view of the sales funnel (including, for example, pending opportunities, billed sales, and expected revenue) to support the development of near-term, quarterly, and rolling-month forecasts.
- *Product visibility.* Trend analysis from a product perspective in terms of the product's standing in its life cycle (for example, growing, stable, or stagnant) and based on sales history.

The users of the current views or applications in the data warehouse include sales planners, business managers, product planners, materials planners, and financial analysts. Currently, more than 900 users access the data warehouse.

In addition to expanding current data, new data types will be added in the following areas during 1996:

- Contract information,
- Customer satisfaction,
- Human resources (for instance, head count),
- Sales opportunities, and
- Competitive intelligence.

New views of the information will emerge. These views include an overall business management support system that scans the data warehouse to provide business managers with both trend and cross-domain analyses. Additional processing and storage capacity for the growing data and application functionalities has been roughly estimated, and it can be supported by adding to the MPP configuration of the NCR 3600 computer. In addition, as table sizes grow, implementation of hierarchies of tables containing pre-summarized data can be used to maintain good query response time. These strategies were part of the original system design and thus far have proven effective.

Lessons Learned

The key aspect to the successful implementation of a data warehouse is organizing applications around a common data model. Time pressures to deliver capability to users gave initial control of the data model to the applications. To respond to the very real needs of their user community, applications developers negotiated independent data feeds and stored the data locally within their application. The result was data redundancy. Once the data responsibility and database administrators were consolidated, the feeds had to be consolidated and common access mechanisms to the data had to be provided. The development organization for implementing a data warehouse should separate data designers and administrators from individuals providing applications or views into the data.

When the project was begun and the computer acquired, the computer's sheer capacity resulted in tremendous pressure to move existing applications onto it. When a decision support application exceeded its current computer capacity, the ideal solution seemed to be to place the application in the data warehouse. The application could then be integrated instead of having to invest in a new computer that would accommodate only one application. Not taken into account was the feasibility of the integration and its overall cost (in which significant progress has been made but at a penalty of almost double the original time estimate).

Summary

To manage an enterprise effectively, information must be available that shows relationships, trends, and correlations among the different functional views of the business. A data warehouse combines large amounts of data from operational and financial systems into a repository that can be accessed by many users having varying needs for information and associated analyses. These needs provide specific challenges in implementing a data warehouse, including dealing with a large volume of data, flexible viewing capabilities, data integrity, and security.

The focus of the WINS project is to implement a technical architecture that will best facilitate moving the business forward as evolution and change continue. User capability is constantly added, as is the corresponding new data. The data warehouse solution for decision support needs provides AT&T-NS with the business management tools necessary to understand and evolve the business more effectively.

***Trademarks**

Business Objects is a trademark of Business Objects S. A.

Ethernet is a registered trademark of Xerox Corp.

Microsoft is a registered trademark and Visual Basic and Windows are trademarks of Microsoft Corp.

PowerBuilder is a trademark of PowerSoft Corp., a wholly owned subsidiary of Sybase Inc.

UNIX is a registered trademark of Novell in the United States and other countries, licensed exclusively through X/Open Company Limited.

Acknowledgments

The original analysis and advocacy for WINS was the work of Peter Miano, Judy Scheffler, and Ed Tovbis. The principles for WINS and many of the concepts in this paper were developed by Jeff Foster. The authors thank Kathy Chin, Steve Kruger, Sai Venkat, Ron Statton, and the entire WINS team for their drive and contributions, as well as Dick Bochan, Cathy Carroll, and the AT&T-NS finance and business process owners for their support of the project.

References

1. K. Orr, "Understanding Data Warehousing," *American Programmer,* Vol. 8, No. 5, May 1995, pp. 2-8.
2. W. H. Inmon, *Building the Data Warehouse,* QED Technical Publications Group, Boston, Massachusetts, 1992.

(Manuscript approved February 1996)

Stacey J. Gelman *is director of manufacturing systems for tł AT&T Network Systems Chief Information Office i Warren, New Jersey. Her prior experience include engineering the process capabilities used on Project Silver Bullet, as well as most recently providing information systems support for sales and marketing. She has a B.A. degree from Queens College in Flushing, New York, and an M.S. degree from the University of Pennsylvania in Philadelphia, both in computer science. She joined AT&T in 1979.*

W. Douglas Peck *is director of common financial application for AT&T Network Systems Chief Information Offic in Whippany, New Jersey. Prior to this, he was responsible for development of a wide variety of operating systems, development tools, and applications, including one of the first mainframe impl mentations of the UNIX* system. Mr. Peck has a B.S. degree in electrical engineering from Brown University in Providence, Rhode Island, and an M.S. degree in computer science from the Stevens Institute of Technology in Hoboken, New Jersey. He joined AT&T in 1969.*