Extended-Enterprise Supply-Chain Management at IBM Personal Systems Group and Other Divisions

Grace Lin, Markus Ettl
Steve Buckley, Sugato Bagchi
IBM T. J. Watson Research Center
Yorktown Heights, New York 10598

David D. Yao
Columbia University
New York, New York 10027

Bret L. Naccarato
IBM Printing Systems Company
Endicott, New York 13760

Rob Allan, Kerry Kim
Lisa Koenig
IBM Personal Systems Group
Research Triangle Park, North Carolina 27709

In 1994, IBM began to reengineer its global supply chain. It wanted to achieve quick responsiveness to customers with minimal inventory. To support this effort, we developed an extended-enterprise supply-chain analysis tool, the Asset Management Tool (AMT). AMT integrates graphical process modeling, analytical performance optimization, simulation, activity-based costing, and enterprise database connectivity into a system that allows quantitative analysis of extended supply chains. IBM has used AMT to study such issues as inventory budgets, turnover objectives, customer-service targets, and new-product introductions. We have implemented it at a number of IBM business units and their channel partners. AMT benefits include over $750 million in material costs and price-protection expenses saved in 1998.

As the world's largest company providing computer hardware, software, and services, IBM makes a wide variety of products, including semiconductors, processors, hard disks, personal computers, printers, workstations, and mainframes. Its manufacturing sites are linked with tens of thousands of suppliers and distribution channels all over the world. A single product line may involve thousands of part numbers with multilevel bills of materials, highly varied lead times and costs, and dozens to hundreds of manufacturing and distribution sites

0092-2102/00/3001/0007/$05.00
1526–551X electronic ISSN
This paper was refereed.

INVENTORY/PRODUCTION—APPLICATIONS
INDUSTRIES—COMPUTERS
MANUFACTURING—SUPPLY CHAIN MANAGEMENT

linked by different transportation modes. Facing the challenges of increasing competition, rapid technology advance, and continued price deflation, the company launched an internal reengineering effort in 1993 to streamline business processes in order to improve the flow of material and information. The reengineering effort focused on improving customer satisfaction and market competitiveness by increasing the speed, reliability, and efficiency with which IBM delivers products to the marketplace.

In 1994, IBM launched an asset-management reengineering initiative as part of the overall reengineering effort. The objectives were to define the supply-chain structure, to set strategic inventory and customer-service targets, to optimize inventory allocation and placement, and to reduce inventory while meeting customer-service targets across the enterprise. The company formed a cross-functional team with representatives from manufacturing, research, finance, marketing, services, and technology. The team identified five areas that needed modeling support for decision making: (1) design of methods for reducing inventory within each business unit; (2) development of alternatives for achieving inventory objectives for senior-management consideration; (3) development and implementation of a consistent process for managing inventory and customer-service targets, including tool deployment, within each business unit; (4) complete evaluation of such assets as service parts, production materials, and finished goods in the global supply network; and (5) evaluation of cross-brand product and unit synergy to improve the management of inventory and risk.

We developed the Asset Management Tool (AMT), a strategic decision-support tool, specifically to address these issues. The integration of AMT with the other asset-management reengineering initiatives has resulted in the successful implementation of extended-enterprise supply-chain management within IBM.

The Asset Management Tool

An extended-enterprise supply chain is a network of interconnected facilities through which an enterprise procures, produces, distributes, and delivers products and services to its customers. As procurement, distribution, and sales have become increasingly global, the supply chains of large companies have become deeply intertwined and interdependent. Today's extended-enterprise supply chains are in fact networks of many supply chains representing the interests of many companies, from supplier's suppliers to customer's customers. Because of this interdependency, a company with an extended supply chain performs well only when it collaborates and cooperates actively with its suppliers and resellers.

A company with an extended supply chain performs well only when it collaborates with its suppliers and resellers.

In high-technology industries, management of the extended-enterprise supply chain becomes very important. At its best, it keeps operating costs low and profits high. But a poorly managed supply chain can reverse that relationship, eroding profits, compromising innovation, and hampering business growth. Early in our efforts, we realized that there were two fundamental keys to overhauling IBM's supply chain. First, we had to reduce and manage uncertainty to promote more accurate forecasts. Second, we had to improve supply-chain flexibility to facilitate quick adaptation to changes in the marketplace. From the outset, we focused on the intrinsic interdependency of an extended-enterprise supply chain. We knew our system would perform as desired only if it reflected the policies and processes used by our suppliers and channels, integrating their value chains with our own. This perspective helped to shape our vision: an integrated modeling and analysis tool for extended-enterprise supply chains. It would be a tool with new methodologies to handle the uncertainties inherent in demand, lead time, supplier reliability, and other factors. It would be scalable, so that it could handle the vast amounts of data describing product structure, supply-chain processes, and component stock information that typify the industry. Finally, the new tool would be equally effective at modeling basic types of supply-chain policies and their interactions, because different companies may use different policies.

We designed AMT to address all of these issues. It is a modeling and analysis system for strategic and tactical supply-chain planning that emerged from various earlier internal IBM reengineering studies [Bagchi et al. 1998; Buckley 1996; Buckley and Smith 1997; Feigin et al. 1996]. It supports advanced modeling, simulation, and optimization capabilities for quantitative analysis of multiechelon inventory systems, along with such features as enterprise database connectivity and internet-based communication. AMT is built on six functional modules: a data-modeling module, a graphical user interface, an experiment manager, an optimization engine, a simulation engine, and a report generator.

The data-modeling module provides a relational data interface, including product structures, lead times, costs, demand forecast and the associated variability information. It has built-in explosion of bills of materials and data-reduction capabilities, and automatic checks for data integrity. It provides access to IBM's global and local operational databases through data bridges.

The graphical user interface (GUI) combines supply-chain modeling with dialog-based entry of supply-chain data. It allows users to build supply networks by dragging and dropping model components, such as manufacturing nodes, distribution centers, and transportation nodes, onto the work space.

The experiment manager facilitates the organization and management of data sets associated with supply-chain experiments. It allows users to view and interactively modify parameters and policies. In addition, it provides automated access to output data generated during experiments and supports a variety of file-management operations.

The optimization engine performs AMT's main function, quantifying the trade-off between customer-service targets and the inventory in the supply network. This module can be accessed from the GUI pull-down menu or called by the simulation engine.

The simulation engine simulates the

performance of the supply chain under various parameters, policies, and network configurations, including the number and location of suppliers, manufacturers, and distribution centers; inventory and manufacturing policies, such as base-stock control, days of supply, build-to-stock, build-to-order, and continuous or periodic replenishment policies. The simulation engine contains an animation module that helps users to visualize the operation of the supply chain or vary parameters and policies while monitoring the simulation output reports.

The report generator offers a comprehensive view of the performance of the supply chain under study, including average cycle times, customer-service levels, shipments, fill rates, and inventory. It also generates financial results, including revenues, inventory capital, raw-material costs, transportation costs, and activity-based costs, such as material handling and manufacturing.

The Optimization Engine

The central function of the optimization engine is to analyze the trade-off between customer-service and inventory investment in an extended-enterprise supply chain. The objective is to determine the safety stock for each product at each location in the supply chain to minimize the investment in total inventory. We view the supply chain as a multiechelon network in which we model each stocking location as a queuing system. In addition to the usual queuing modeling, we incorporated into the model an inventory-control policy: the base-stock control, with the base-stock levels being decision variables. To numerically evaluate such a network, we developed an approach based on decomposition. The key idea is to analyze each stocking location in the network individually and to capture the interactions among different stocking locations through their so-called actual lead times.

We modeled each stocking location as a queue with batch Poisson arrivals and infinite servers with service times following general distributions, denoted as $M^X/G/\infty$ in queueing notation. To do this, we had to specify the arrival and the service processes. We obtained the arrival process at each location by applying the standard MRP demand explosion technique to the production structure. The batch Poisson arrival process has three main parameters: the arrival rate, and the mean and the variance of the batch size. It thus accommodates many forms of demand data; for instance, demand in a certain period can be characterized by its minimum, maximum, and most likely value. The service time is the actual lead time at each stocking location. The actual lead time at a stocking location can be derived from its nominal lead time (for example, the manufacturing or transportation time) along with the fill rate of its suppliers. In particular, when a supplier has a stock-out, we have to add the resulting delay to the actual lead time. This delay is the time the supplier takes to produce the next unit to supply the order. In our model, we derive the additional delay from Markov-chain analysis.

AMT embodies a creative coupling of optimization, performance evaluation, and simulation.

With the arrival and service processes in place, we can analyze the queue and derive performance measures, such as inventory, back-orders, fill rates, and customer-service levels. The key quantity in the analysis of a stocking location, i, is the number of jobs in the $M^X/G/\infty$ queue, denoted N_i, which can be derived from standard queueing results [Liu, Kashyap, and Templeton 1990]. To alleviate the computational burden in large-scale applications, we approximated N_i by a normal distribution. This way, we need to derive only the mean and the variance of N_i, both of which depend on the actual lead time, which is the service time in the queuing model. Figure 1 shows a snapshot of the dynamics at a stocking location.

The objective of the optimization model is to minimize the total expected inventory capital in the supply network. This total is a summation over all stocking locations, each of which carries two types of inventory: finished goods (on-hand) inventory, and work-in-process (on-order) inventory. The constraints of the optimization model are the required customer-service targets. They are represented as the probability, say 95 or 99 percent, that customer orders are filled by a given due date. Our formulation allows users to specify customer-service targets separately for each demand stream. We first derive the fill rates for each end product to meet the required customer-service target. These fill rates relate to the actual lead times of all upstream stocking locations, via the bills-of-materials structure of the network, and to the actual lead times. The model thus captures the interdependence of different stocking locations, in particular the effect of base-stock levels and fill rates on customer-service. Related models in supply-chain and distribution networks include those of Lee and Billington [1993], Arntzen et al. [1995], Camm et al. [1997],

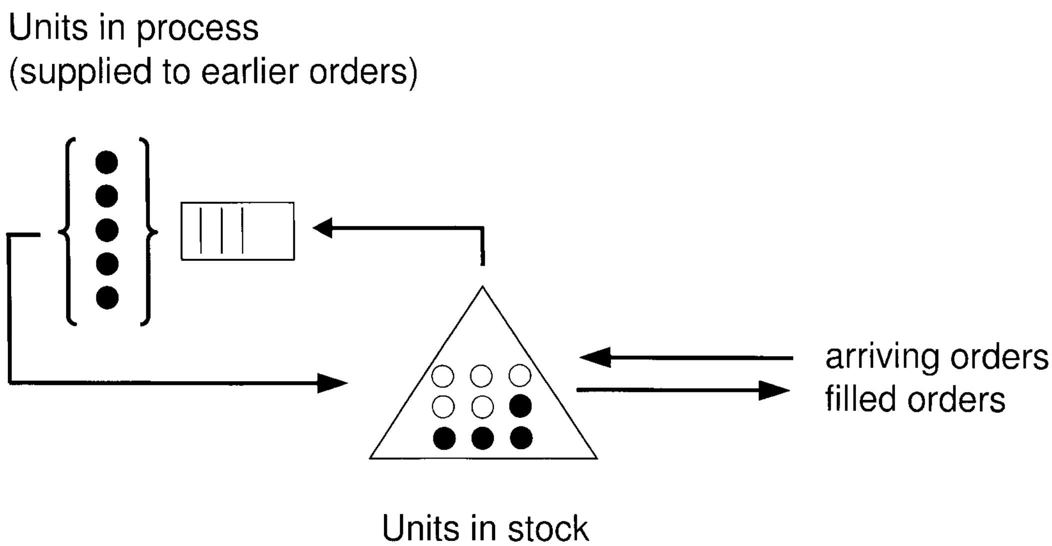

Figure 1: In this snapshot of the system dynamics at a stocking location, the base-stock level is nine, and when there are four units in stock, the other five units have been supplied to earlier orders, which translates into the five jobs in process.

Kruger [1997], Graves, Kletter, and Hetzel [1998], and Andersson, Axsaeter, and Marklund [1998].

To allow fast execution of the optimization, we derived analytical gradient estimates in closed form and implemented a gradient search algorithm to generate optimal solutions. Technical details of this work are presented by Ettl et al. [1998] and in the Appendix. In addition to the gradient search, we developed a heuristic optimization procedure based on product clustering. To validate the solution approach, we compared it against exhaustive searches for test problems of moderate size. For large-scale, industry-size applications, the model has been extensively tested at several IBM business units.

The Simulation Engine

The simulation engine allows users to simulate various supply-chain policies and in particular to verify and fine-tune the performance of the solutions generated by the optimization engine. We built the simulation engine upon SimProcess [Swegles 1997], a general-purpose business-process simulator that was developed jointly by IBM Research and CACI Products Company. The simulation engine preserves the capabilities of SimProcess while adding a supply-chain modeling functionality. Specifically, it provides modeling functions for the following supply-chain processes:

—The customer process represents outside customers that issue orders to the supply chain, based on the modeled customer demand. It can also model information about the desired customer-service target and priority for the customer.

—The manufacturing process models assembly processes, buffer policies, and replenishment policies. It can also be used to model suppliers.

—The distribution process models distribution centers and can also be used to model retail stores.

—The transportation process models transportation time, vehicle loading, and transportation costs.

—The forecasting process represents product forecasts, including promotional and stochastic demand, for future periods.

—The inventory-planning process models periodic setting of inventory target levels. Underlying this process is the AMT optimization engine that computes recommended inventory levels at the various stocking locations in a supply chain based on desired customer-service target.

The simulation engine allows the user to vary a set of input parameters while monitoring output reports to obtain the best set of output values. All input and output parameters reside in the AMT modeling database. Users provide input parameters for the simulation in the form of random variables with stochastic distributions; these include manufacturing lead times, transportation times, material-handling delay times, demand forecasts, product quantity required in a bill of material, and supplier reliability. The stochastic distribution functions supported include beta, Erlang, exponential, gamma, normal, lognormal, Poisson, triangular, uniform, Weibull, and user-defined distributions.

We designed the simulation engine to enable scenario-based analyses in which supply-chain parameters, such as the number and location of suppliers, manufacturers, and distribution centers, inventory levels, and manufacturing, replenishment, and transportation policies (build-to-plan, build-to-order, assemble-to-order, continuous replenishment, periodic replenishment, full truckload, less-than-truckload, and so forth) are varied across simulation runs. For each simulation run, the user can specify a planning horizon, the number of replicating scenarios (sample runs), and a warm-up period during which statistics are not retained. The length of the planning horizon depends on the particular application in question and the availability of historical demand forecasts. We typically choose a horizon that is between six and 12 months.

The simulation-run outcome is in the form of measurement reports that can be generated for turnaround times, customer-service, fill rates, stock-out rates, shipments, revenue, safety stock, and work-in-process. To analyze financial impacts, users can employ the following items, all of which are monitored during the simulation: cost of raw material; revenue from goods sold; activity-based costs, such as material handling and manufacturing; inventory-holding costs; transportation costs; penalties for incorrectly filled or late orders delivered to customers; credits for incorrectly filled or late deliveries from suppliers; cost of goods returned by customers; and credits for goods returned to suppliers.

System Integration and Technical Innovations

We integrated the six functional modules of AMT in a system architecture that is flexible enough to accommodate users' varying computational needs. The architecture is based on a client-server programming model in which one can conduct experiments using the resources of a computer network (Figure 2). The AMT client side provides a set of functions for viewing the graphical user interface and dialog-based data entry. The AMT server side, which typically resides on a powerful workstation or midrange computer, provides the full modeling and analysis functionality. For users with access to low-powered computers, such as laptops, we developed an architecture in which the AMT client side is implemented as a platform-independent Java application or applet; web-enabled clients allow users to access AMT through a web browser.

To manage supply-chain operations, AMT requires data about the different stages and processes that products go through. This data is accessible through a relational modeling database that is connected to the server through a relational interface. The database stores the information associated with the various modeling scenarios, including the supply-chain structure, product structure, manufacturing data, and demand forecasts. The product structures are derived from a top-down bills-of-materials explosion that is processed for each end product. We extracted all product data from corporate databases and from local site data sources.

To facilitate data extraction, we developed a number of database connectivity modules that provide automated database access, extract production data, and feed them into the modeling database. All connectivity modules have built-in bills-of-materials explosion functionality. To detect inconsistencies in data recording caused by missing or incomplete informa-

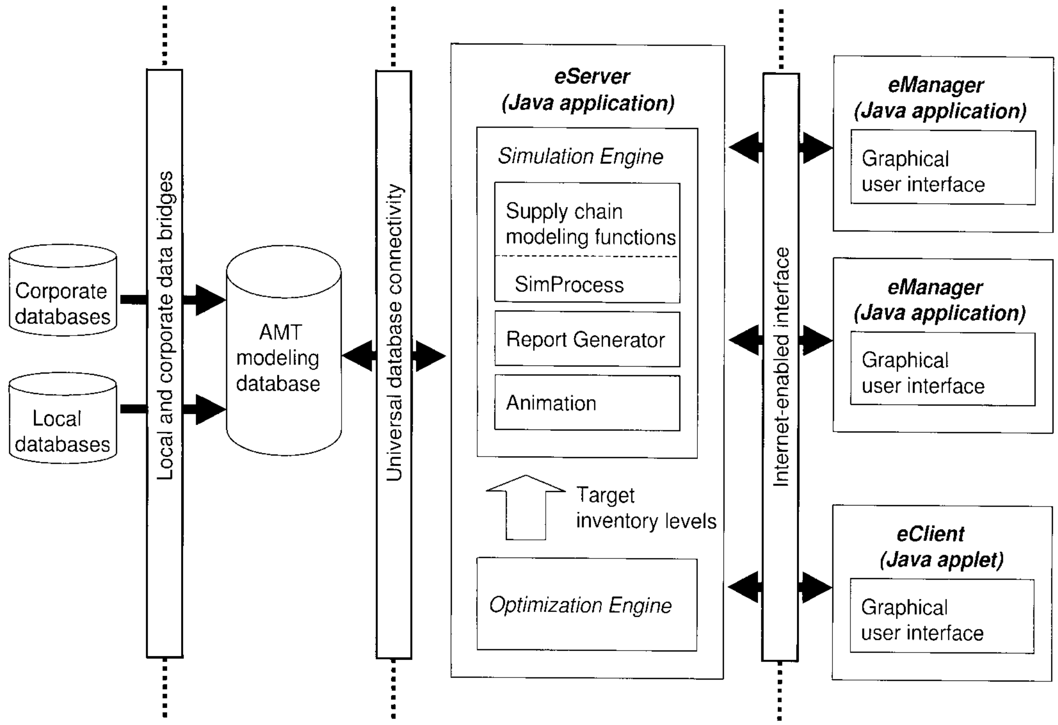

Figure 2: AMT is implemented using a client-server architecture in which the modeling functionality is separated from the graphical user interface. The modeling engines reside on a server computer (eServer). The graphical user interfaces are piped to client computers that are implemented as either Java applications (eManager) or Java applets (eClient). The AMT modeling database can be accessed through a relational database interface. It contains such supply chain data as bills of materials, demand forecasts, lead times, costs, inventory policies, and customer-service requirements. Local and corporate data bridges provide automated access to enterprise data sources.

tion pertaining to the bills of materials, we added database consistency checks that generate missing data reports and reduce the data set to a consistent level that can be downloaded to the modeling database. The data-collection process allows the user to supply missing data in relational tables that can be merged with the output of the explosion. To keep the complexity of the bills-of-materials explosion manageable, we implemented data-reduction routines through which one can eliminate noncritical components automatically, based on the item's value class or annual requirements cost.

AMT's graphical user interface allows modelers to build supply networks for a variety of supply chains by dragging and dropping generic supply-chain components on the workspace (Figure 3). Sophisticated algorithms are encapsulated in the components. For instance, clicking the "PSG manufacturing" node will bring up screens for the user to specify parameters and policies, such as delay time, manufacturing lead times, bills of materials, and such manufacturing policies as build to order or build to plan. AMT also supports

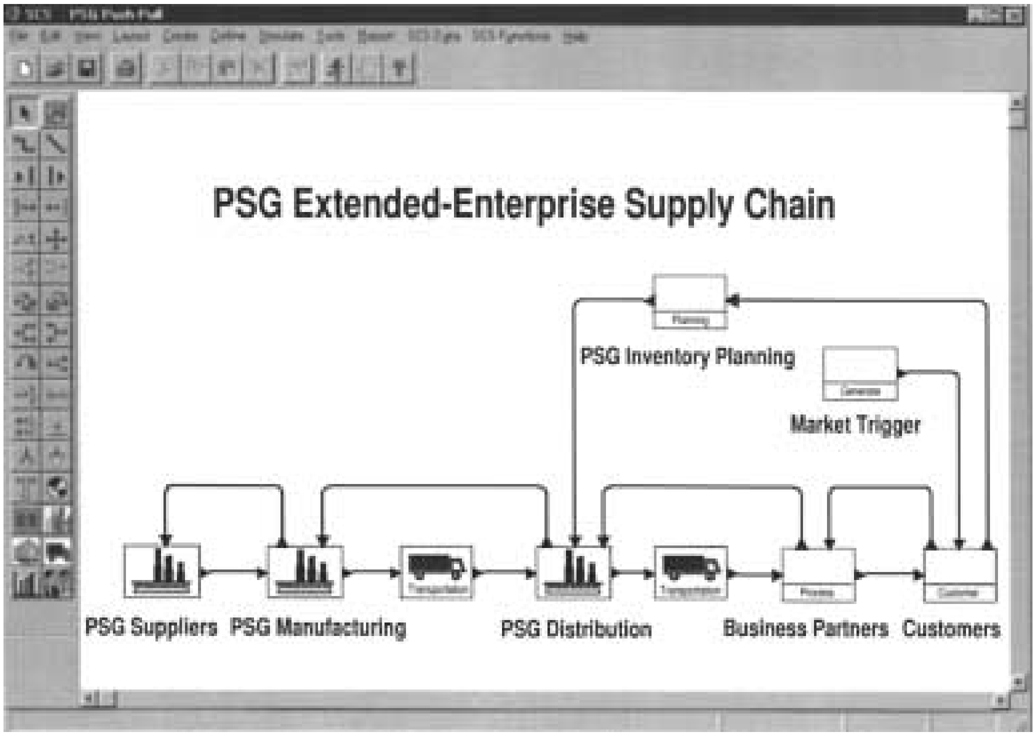

Figure 3: AMT provides a graphical user interface that allows one to interactively construct supply chain scenarios. In this example of an extended-enterprise supply chain, business partners (PSG Business Partners) send orders to a distribution center (PSG Distribution). The distribution center processes the orders and sends products to a transportation node that ships the products to the business partners. The distribution center needs to replenish its stock from time to time, so that it sends replenishment orders to the manufacturing site (PSG Manufacturing) that assembles finished products. The manufacturing site in turn replenishes its parts supply by sending orders to its suppliers (PSG Suppliers). An inventory-planning node (PSG Inventory Planning) representing the AMT optimization engine computes optimal inventory levels for the distribution center based on forecasts of customer demand.

hierarchical process modeling. The user can drill down to include other layers of the supply chain, adding scalability to the modeling approach. The customer node captures demand, forecast, and customer-service requirements. We built in animation to help users visualize the supply-chain activities of orders, goods, and trucks moving between nodes. As the simulation is running, users can view reports, such as service or inventory reports, to see the current status of the simulation. In addition to these real-time reports, AMT also offers the financial and performance reports that we discussed earlier.

An important feature of AMT is the complementary functionality of the optimization and simulation engines. With the optimization engine, the user can perform fast yet very deep what-if analyses, which are beyond the capability of any standard simulation tool. With the simulation engine, the user can invoke the inventory module to perform periodic recalculations of optimal inventory levels while simulating dynamic supply-chain processes and policies. The user can run simulations on optimized solutions, observing how different supply-chain policies at different locations affect the supply chain's performance. Simulation results can also be used to adjust parameters of the simulation or optimization runs. An automated interface between the simulation engine and the optimization engine allows users to invoke optimization periodically during a simulation run, for example to recalculate target inventory levels according to the latest forecast of demand. Users can also use the optimization engine to periodically generate build plans in a mixed push-pull manufacturing environment, taking into account service targets and system uncertainty.

In summary, AMT embodies a creative coupling of optimization, performance evaluation, and simulation, integrated with data connectivity and an Internet-enabled modeling framework. This makes it a powerful and versatile tool for capturing the stochastic and dynamic environment in large-scale industrial supply chains. We model extended-enterprise supply chains as networks of inventory queues, using a decomposition scheme and queuing analysis to capture the performance of each stocking location. We developed multiechelon, constrained inventory-optimization algorithms that use conjugate gradient and heuristic searches for efficient large-scale applications. We developed a supply-chain simulation library consisting of an extensive set of supply-chain processes and policies for modeling various supply-chain environments with little programming effort. It offers performance measures, financial reports, and activity-based costing down to the level of individual stock-keeping units. It also gives the user a way to validate and fine-tune supply-chain parameters based on analytical results.

Extended-Enterprise Supply Chain Management at IBM Personal Systems Group

The IBM Personal Systems Group (PSG) is responsible for the development, manufacture, sale, and service of personal computers (for example, commercial desktops, consumer desktops, mobile products, workstations, PC servers, network PCs, and related peripherals). PSG employs over 18,500 workers worldwide. Sales and marketing groups are located in major metropolitan areas, with manufacturing plants located in the United States, Latin America, Europe, and Asia. In 1998, PSG sold approximately 7.7 million computers under such brand names as IBM PC, Aptiva, ThinkPad, IntelliStation, Netfinity, and Network Station.

Increased competition from such PC manufacturers as Dell and Gateway, which use a direct, build-to-order business model, prompted PSG in 1997 to reevaluate its business practices and its relationships with its supply-chain partners. The goal was to design and implement a hybrid business model, one that incorporated the best features of the direct model (build to order, custom configuration, and inventory minimization) and the best features of the indirect model (final configuration, high customer service, and support), selling products through multiple channels.

PSG formed a cross-functional team in April 1997 with the task of quantifying the relationship between customer service and inventory throughout the extended supply chain under the existing business model and under various proposed channel-assembly alternatives. We used production data from a subset of PSG's commercial desktop products to develop a baseline supply-chain model in AMT. The model was triggered by end-user demand, reseller ordering behavior, IBM manufacturing and inventory policies, supplier performance, and lead-time variability. We collected actual end-user sales data for 22 reseller locations over five months. Resellers' ordering behavior was influenced by many factors, such as gaming strategies, marketing incentives, confidence in supplier reliability, and stocking for large customer purchases. Modeling each individual activity would have been too complex. Our model captured the aggregate ordering for each PSG reseller by substituting alternative ordering policies, representing current levels of sales activity in the channel. For example, if a particular reseller held an average of 60 days of inventory, the model established a target base-stock level representing 60 days of channel inventory for this reseller. To see what would happen if resellers changed their ordering policies, we changed the levels of channel inventory in the AMT model and ran different what-if scenarios. For each ordering policy, we assumed that a reseller would stock a product at a given level of days of supply.

During the normal course of business, PSG forecasts its manufacturing volumes over a rolling 13-week horizon. The current week's forecast becomes the build plan, which then pushes products built at PSG's manufacturing sites to the distribution warehouse where they are held until the products are eventually ordered, or pulled, by a reseller. This type of replenishment policy captured the logic of PSG's hybrid push-pull manufacturing and ordering system in which PSG built products to a forecast and held them as finished goods in the warehouse until it received orders from its resellers. This system is not a true pull system because

PSG's channel look-back expenses dropped by more than $100 million.

product availability influences reseller ordering. Likewise, the system is not a true push system because the backlog of resellers' orders influences the schedules at PSG manufacturing sites. To effectively capture variability caused by component shortages, capacity constraints, and requirements for minimum lot sizes, we analyzed the range of the 13-week forecasts.

PSG set a service target for customer deliveries of three days, 95 percent of the time, which translated directly into the customer-service constraint required by the AMT optimization engine. Combining the simulation engine with the optimization engine, the model recalculated the base-stock levels every week, according to the latest available forecast of demand so that customer orders could be filled within three days 95 percent of the time. This replenishment policy formed the basis for PSG's supplier orders for components and subassemblies and for its subsequent manufacturing activity. In Phase 1 of the project, we used a reduced data set to construct a simplified prototype model of PSG's supply chain to test assumptions, to investigate alternative modeling algorithms, and to better understand possible limitations of the AMT application.

In Phase 2 of the project, we developed more detailed modeling scenarios to vary channel inventory and to incorporate a channel-assembly policy at the resellers. PSG delivers to its resellers two types of products, (1) standard machine-type models (MTMs), which are fully configured and tested computers, and (2) so-called open-bay machines, which are nonfunctional, basic computers without such preconfigured components as memory, hard files, and CD ROMs. These open bays allow resellers to assemble machines according to specific customer requirements. We found that some resellers converted open bays into standard MTMs as needed and then sold them to their customers. We refer to this as an example of flexibility because resellers can use their current open-bay inventory to fill orders for standard MTMs, instead of stocking open bays exclusively to fill orders for nonstandard MTMs. Other resellers stockpiled open-bay inventory, and if they needed standard MTMs to fill an order, they would reorder from PSG instead of configuring an open bay already in stock (an example of inflexibility). Both methods affect inventory and customer service. Because reseller flexibility could not be defined accurately, we designed different sets of simulation experiments with the intent to bound, or frame, the true impact of channel assembly within the two extreme cases of 100 percent reseller flexibility and 100 percent reseller inflexibility.

We validated the accuracy of the AMT models by comparing the outputs of the simulation runs to historical PSG data. We adjusted our modeling assumptions and parameters as necessary and ran multiple simulations using different parameters and policies. The key results of the study can be summarized as follows:

—Implementing channel assembly based on PSG's existing product structure, low volume environment, and present supply-chain policy reduces inventory very little (inflexible reseller channel behavior).

—Allowing resellers to configure any MTM from their stock of components could improve customer service by two percent and simultaneously reduce inventories by 12 percent (flexible reseller channel behavior).

—Consolidating the demand at 22 configuration sites into three large hubs could improve customer service by six percent and reduce inventories by five percent.

—Based on the existing push-pull supply-chain policy, PSG can reduce channel inventory by 50 percent without affecting its customer-service level. The overall supply-chain inventory levels were far in excess of the optimum needed to maintain PSG's service target.

This and subsequent projects brought together four functional groups—marketing and sales, manufacturing, distribution, and development—to seek a company-wide consensus on PSG's strategic direction and subsequent actions. Our studies contributed directly to PSG's advanced fulfillment initiative (AFI), an effort to increase flexibility in the reseller channel by improving parts commonality in PSG's product structure [Narisetti 1998]. Also, PSG management endorsed the reduction of the number of configuration sites, as a result of changing channel price-protection terms and conditions. The specific terms and conditions were tied to the output of the AMT model, and they were implemented in November 1997 after a series of related enhancements to the logistics process.

PSG has based many of its decisions on how to prioritize project deployment and manage channel inventory on the results of subsequent AMT analyses. While the analysis that drove PSG's initial business transformation was conducted in 1997, the 1998 business benefits were substantial.

The more accurate a reseller's forecasts, the higher the level of service.

PSG reduced its overall inventory by over 50 percent from year-end 1997 to year-end 1998. As a direct consequence of this inventory reduction, PSG's channel look-back expenses dropped by more than $100 million from 1997 levels. Look-back expenses account for payments to distributors and business partners that compensate for price actions on the inventory they are holding. In addition, by selling products four to six weeks closer to when the components are procured, PSG saved an additional five to seven percent on product cost. This equates to more than $650 million of annual savings.

In the months following the original assessment, we conducted further supply-chain studies, including analyses that (1) incorporated the supply chains of business partners; (2) modeled additional geographies; (3) assessed the impact on inventory and customer service of delaying final assembly to the reseller's distribution facilities; and (4) estimated the impact on inventory of reducing manufacturing cycle times. These studies have helped PSG's business partners make more informed decisions on supply-chain policy. In particular, they have led IBM and its major business partners to establish a colocation policy. In colocation, a business partner locates its distribution space inside of IBM, eliminating the need for costly handling and transportation among different sites. Finally, because we found that forecast accuracy greatly affected inventory and customer service, PSG used the AMT to determine the level of service it would promise to its business partners, based on their ability to provide accurate forecasts. The more accurate a reseller's forecasts, the higher the level of service PSG would provide to that reseller. This policy is unprecedented in the industry and has been favorably received by PSG's business partners. Overall, PSG believes that the AMT has been an invaluable asset in developing and implementing world-class supply-chain-management policies.

Other AMT Applications Across IBM

AMT has also been applied and deployed in other IBM manufacturing divisions, including the printing systems division (PSC), the midrange computer division (AS/400), the office workstation division (RS/6000), the storage systems division (SSD), the mainframe computer division (S/390), and PSG's European market. A number of PSG's business partners have used AMT, including Pinacor, GE Capital, and Best Buy. IBM's Industry Solution Unit uses the tool externally for consulting engagements. Following are brief descriptions of three recent AMT engagements:

The IBM Printing Systems Company (PSC) is a leading supplier of printer solutions for business enterprises. The product line includes printers for office printing to high-volume production printing. The company employs approximately 4,550 people, with total gross revenue for 1998 of $1.95 billion. In 1996, PSC conducted an intensive testing process on the AMT over a five-month period. In its assessment report, the testing team concluded that AMT produces accurate results, provides productivity improvements over existing supply-chain-management and inventory tools, and improves PSC's precision in validating and creating inventory budgets and turnover objectives. PSC then used AMT to study the effects of forecast accuracy, product structure, the introduction of a new distribution center, and different business scenarios on the performance of the supply network for different product families. In one of the cases alone, it reported inventory savings of $1.6 million, which represented 30 percent of the total inventory holding cost.

Financial savings amount to more than $750 million at PSG in 1998.

IBM's AS/400 division manufactures midrange business computers and servers, providing more than 150 models and upgrades with up to 1,000 features. Assembling these systems requires several thousand unique part numbers, approximately 1,000 of them used at the highest level of assembly just prior to building a complete system. Providing customers with the flexibility to customize the equipment they order by selecting features creates manufacturing complexity and efficiency challenges. The division used AMT to analyze and quantify the impact on inventory and on-time delivery of feature reduction, feature substitution, parts commonality, and delayed customization. The analysis showed that eliminating low-volume parts would improve inventory turnover by 15 percent and that substituting and postponing their final assembly would improve inventory turnover by approximately 30 percent. The AS/400 division has reduced its feature count by approximately 30 percent since 1998 with steady growth in total revenue.

In 1995, IBM established a quick-response service program to provide rapid delivery for customers buying selected mid-range computer memory, storage, and features. In September 1998, IBM instituted the quick-response program as a front end to provide real e-commerce for our large business partners. IBM used AMT to analyze the trade-off between service and inventory in choosing an optimum performance point. It later used it to assess the impact of the quick-response program on allocating inventory between manufacturing and distribution centers. The results helped IBM to maximize business efficiency and contributed to doubling the growth of quick-response revenue in 1998.

Conclusions

The AMT effort uses advanced OR techniques and combines technical innovations with practical and strategic implementations to achieve significant business impacts. IBM has used AMT to address a wide range of business issues, including inventory management, supply-chain configuration, product structure, and replenishment policies. AMT has been implemented in a number of IBM business units and their business partners. Financial savings through the AMT implementations amount to more than $750 million at PSG in 1998 alone. Furthermore, AMT has helped IBM's business partners to meet their customers' requirements with much lower inventory and has led to a co-location policy with many business partners. It has become the foundation for a number of supply-chain-reengineering initiatives. Several IBM business partners view the AMT analyses as key milestones in their collaboration with IBM in optimizing the extended-enterprise supply chain.

Acknowledgments

We gratefully acknowledge the contributions and support of the following people: Chae An, Ray Bessette, Rick Bloyd, Harold Blake, Richard Breitwieser, J. P. Briant, Bob Chen, Feng Cheng, Arthur Ciccolo, Daniel Connors, Ian Crawford, Anthony Cyplik, John Eagen, Brian Eck, Gerry Feigin, Angela Gisonni, John Konopka, Tom Leiser, Tony Levas, Nikorn Limcharoen, Joe Magliula, Barbara Martin, Larry McLaughlin, Bob Moffat, Gerry Murnin, Nitin Nayak, Jim Nugent, Lynn Odean, Krystal Reynolds, Richard Shore, Mukundan Srinivasan, Jayashankar Swaminathan, David Thomas, Bill Tulskie, Burnie Walling, Wen-Li Wang, and James Yeh.

APPENDIX

Optimization of Multi-Echelon Supply Networks with Base-Stock Control

Here we provide a brief overview of the key points of the mathematical model in the optimization engine. Ettl, Feigin, Lin and Yao [1998] give the full details, including topics that we do not touch upon here, such as the treatment of nonstationary demands, the related rolling-horizon implementation, the derivation of the gradients, and many preprocessing and postprocessing steps.

We specify the configuration of the supply network using the bills-of-materials structure of the products. Each site in the network is either a plant or a distribution center. Associated with each site and each product processed at the site is a multilevel bill of material. Each site has storage areas, which we refer to as stores, to hold both components that appear on the bills of materials and finished products, which correspond to input stores and output stores. The subscripts i and j index the stores, and S denotes the set of all stores in the network. We assume a distributed inventory-control mechanism whereby each store follows a base-stock control policy for managing inventory. The policy works as follows: When the inventory position (that is, on-hand plus on-order minus backorder) at store i falls below some specified base-stock level, R_i, a site places a replenishment order. In our model, R_i is a decision variable.

For each store i, there is a nominal lead time, L_i, with a given distribution. The nominal lead time corresponds to the production time or transshipment time at the site where the store resides, assuming there is no delay (due to stock-out) in any upstream output stores. The actual lead time, $\tilde{L}_i$, in contrast, takes into account possible additional delays due to stock-out. Whereas L_i's are given data, $\tilde{L}_i$'s are derived performance measures.

To analyze the performance of each store i, we use an inventory-queue model, for example, Buzacott and Shanthikumar's [1993]. Specifically, we combine the base-stock control policy with an $M^X/G/\infty$ queue model, where arrivals follow a Poisson process with rate λ_i, and each arrival brings in a batch of X_i units, or orders. The batch Poisson arrival process is a good trade-off between generality and tractability. In particular, it offers at least three parameters to model the demand data: the arrival rate and the first two moments of the batch size (whereas a simple Poisson arrival process has only one parameter).

To derive the performance measures at each store i, we need to first generate the input process to the $M^X/G/\infty$ queue. To do this, we take the demand stream (forecast or real) associated with each class, translate it into the demand process at each store by going through the bills-of-materials structure level by level, and shift the time index by the lead times at each level. This process is quite similar to the explosion and offsetting steps in standard MRP analysis. A second piece of data needed for the $M^X/G/\infty$ queue is the service time, which we model as the actual lead time.

Let N_i be the total number of jobs in the queue $M^X/G/\infty$ in equilibrium. Following standard queueing results [Liu, Kashyap, and Templeton 1990], we can derive the mean and the variance of N_i, denoted as μ_i and σ_i^2. We then approximate N_i with a normal distribution:

$$N_i = \mu_i + \sigma_i Z, \tag{1}$$

where Z denotes the standard normal variate. Accordingly, we write the base-stock level as follows:

$$R_i = \mu_i + k_i\sigma_i, \tag{2}$$

where k_i is the so-called *safety factor*. As R_i and k_i relate to each other via the above relation, either can serve as the decision variable. Let I_i be the level of on-hand inventory, and B_i the number of back-orders at store i. These relate to N_i and R_i as follows:

$$I_i = [R_i - N_i]^+ \quad \text{and } B_i = [N_i - R_i]^+, \tag{3}$$

where $[x]^+ = \max(x,0)$. We can then derive the expectations:

$$E[I_i] = \sigma_i H(k_i), \text{ and } E[B_i] = \sigma_i G(k_i) \tag{4}$$

where

$$H(k_i) = \int_{ki}^{\infty} (k_i - z)\phi(z)dz \text{ and } G(k_i) = \int_{ki}^{\infty} (z - k_i)\phi(z)dz, \tag{5}$$

and $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ is the density function of Z. Furthermore, writing $\Phi(x) = \int_0^x \phi(z)dz$, the distribution function of Z, and $\bar{\Phi}(x) = 1 - \Phi(x)$, we can derive the stock-out probability p_i and the fill rate f_i at store i as follows:

$$p_i = \bar{\Phi}(k_i), \text{ and } f_i = 1 - \sigma_i\phi(k_i)/\mu_i - \bar{\Phi}(k_i). \tag{6}$$

All of the above performance measures involve the actual lead time at store i, which can be expressed as follows:

$$\tilde{L}_i = L_i + \max_{j \in S>i}(\tau_j), \tag{7}$$

where $S_{>i}$ denotes the set of stores that supply the components needed to build the units in store i, and τ_j denotes the additional delay at store $j \in S_{>i}$. As τ_j is quite intractable in general, with queueing analysis, we have derived the following approximation:

$$\tau_j = \tilde{L}_j r_j \text{ where } r_j := \frac{E[B_j]}{p_j(R_j + 1)}. \tag{8}$$

Intuitively, $E[B_j]/p_j$ is the average number of back-orders at location j conditioned upon a stock-out there, and each of these back-orders requires an average time of $\tilde{L}_j/(R_j + 1)$ to fill, that is, during the stock-out, on the average, there are $(R_j + 1)$ outstanding orders in process.

Customer demands are supplied from a set of end stores, S_0, stores at the boundary of the network. Consider a particular customer class, and suppose its demand is supplied by one of the end stores, $i \in S_0$. Let W_i denote the waiting time to receive an order. The required customer-service target is

$$P[W_i \leqslant \beta_i] \geqslant \alpha_i, \tag{9}$$

where β_i and α_i are given data. When the demand is supplied from on-hand inventory, the delay is simply the transportation time T_i, time to deliver the finished products to customers, which is given; otherwise, there is an additional delay of τ_j. Hence,

$$P[W_i \leqslant \beta_i] = f_i P[T_i \leqslant \beta_i] + (1 - f_i) P[T_i + \tau_i \leqslant \beta_i].$$

For the above to be at least α_i, we need to set f_i, the fill rate, to the following level:

$$f_i = \frac{\alpha_i - P[T_i + \tau_i \leqslant \beta_i]}{P[T_i \leqslant \beta_i] - P[T_i + \tau_i \leqslant \beta_i]}. \tag{10}$$

The quantity τ_i involved in the right-hand side of the above equation can be expressed as $\tau_i = \tilde{L}_i\, r_i$, following (8). Since r_i involves B_i and R_i, both of which are functions of k_i, and so is f_i, we need to solve a fixed-point problem defined by the equation in (8) to get f_i (or k_i). In the iterations involved in the optimization procedure, however, this fixed-point problem can be avoided by simply using the r_i value obtained from the previous iteration. Once we derive f_i and k_i, the base-stock level (2) and the stock-out probability (6) then follow.

The objective of our optimization model is to minimize the total expected inventory capital throughout the supply network while satisfying customer-service requirements. Each store has two types of inventory: on-hand inventory and work-in-process (WIP) inventory. (The WIP includes the orders in transition, that is, orders being transported from one store to another.) From the above discussion, the expected on-hand inventory at store i is $E[I_i] = \sigma_i H(k_i)$, and the expected WIP is $E[N_i] = \mu_i$. Therefore, the objective function takes the following form:

$$C(\mathbf{k}) = \sum_{i \in S} [c_i' \mu_i + c_i\, \sigma_i H(k_i)], \tag{11}$$

where c'_i and c_i denote the inventory capital per unit of the on-hand and WIP inventory, respectively, with c_i assumed given, and c'_i derived from the c_i's along with the BOM. We want to minimize $C(\mathbf{k})$, subject to meeting the fill-rate requirements in (10), for all the end stores: $i \in S_0 \subset S$. This is a constrained nonlinear optimization problem. We derive the partial derivatives $\partial/\partial k_j\, C(k)$, all in explicit analytical forms based on the relations derived above (and others). We use these in a conjugate-gradient search routine, for example that of Press et al. [1994]. As the surface of the objective function is quite rugged, to avoid local optima, we also implemented several heuristic search procedures. For instance, evaluate a set of randomly generated initial points and pick the best one (in terms of the objective value) to start the gradient search.

References

Andersson, J.; Axsaeter, S.; and Marklund, J. 1998, "Decentralized multi-echelon inventory control," *Production and Operations Management*, Vol. 7, No. 4, pp. 370–386.

Arntzen, B. C.; Brown, G. G.; Harrison, T. P.; and Trafton, L. L. 1995, "Global supply chain management at Digital Equipment Corporation," *Interfaces*, Vol. 25, No. 1, pp. 69–93.

Bagchi, S.; Buckley, S.; Ettl, M.; and Lin, G. 1998, "Experience using the supply chain simulator," *Proceedings of the Winter Simulation Conference*, Washington, DC, December, pp. 1387–1394.

Buckley, S. 1996, "Supply chain modeling," *Proceedings of the Autofact Conference*, Detroit, Michigan, pp. 749–756.

Buckley, S. and Smith, J. 1997, "Supply Chain Simulation," *Georgia Tech Logistics Short Course*, Atlanta, Georgia, pp. 1–17.

Buzacott, J. A. and Shanthikumar, J. G. 1993, *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, New Jersey.

Camm, J. D.; Chorman, T. E.; Dill, F. A.; Evans, J. R.; Sweeney, D. J.; and Wegryn, G. W. 1997, "Blending OR/MS, judgment, and GIS: Restructuring P&G's supply chain," *Interfaces*, Vol. 27, No. 1, pp. 128–142.

Ettl, M.; Feigin, G.; Lin, G.; and Yao, D. D. 1998, "A supply network model with base-stock control and service requirements," *Operations Research*, forthcoming.

Feigin, G.; An, C.; Connors, D.; and Crawford, I. 1996, "Shape up, ship out," *ORMS Today*, Vol. 23, No. 2, pp. 1–5.

Graves, S.; Kletter, D. B.; and Hetzel, W. B. 1998, "A dynamic model for requirements planning with application to supply chain optimization," *Operations Research*, Vol. 46, No. 3, pp. S35–S49.

Kruger, G. A. 1997. "The supply chain approach to planning and procurement management," *Hewlett-Packard Journal*, Vol. 48, No. 1, pp. 1–9.

Lee, H. L. and Billington, C. 1993, "Material management in decentralized supply chains," *Operations Research*, Vol. 41, No. 5, pp. 835–847.

Liu, L.; Kashyap, B. R. K.; and Templeton, J. G. C. 1990, "On the $GI^X/G/\infty$ system," *Journal of Applied Probability*, Vol. 27, No. 3, pp. 671–683.

Narisetti, R. 1998, "How IBM turned around its ailing PC Division," *Wall Street Journal*, October 3, p. B1.

Press, W. H.; Teukolsky, S. A.; Vettering, W. T.; and Flannery, B. P. 1994, *Numerical Recipes in C*, second edition, Cambridge University Press, New York.

Swegles, S. 1997, "Business process modeling with SIMPROCESS," *Proceedings of the Winter Simulation Conference*, Piscataway, New Jersey, pp. 606–610.

Bob Moffat, general manager for manufacturing, procurement, and fulfillment at IBM Personal Systems Group, said during the presentation of the paper at the Edelman competition: "We reduced our channel inventory from over three months to approximately one month. As a direct consequence of this inventory reduction, our division has reduced 1998 price protection expenses by over $100M from the previous year. Price protection expenses are what we reimburse business partners whenever we take a price action on products they are holding. We had reduced our end-to-end inventory from four and a half months to less than two months by the end of 1998. By closing the gap between component procurement and product sale by four to six weeks, there is a savings on product cost of at least five percent. This equates to more than $650 million of annual savings. AMT has improved our relationships with business partners, making them more efficient, more productive, and ultimately more powerful in the marketplace. I believe this will lead to a fundamental change in our business culture, a unification of basic value among suppliers, manufacturers, and resellers."

Jean-Pierre Briant, IBM vice president for integrated supply chain, further explained: "The AMT tool has found application in almost every supply chain within IBM. It helps us understand our extended supply chain—from our suppliers' suppliers to our customers' customers. We have deployed the AMT tool to assist external companies in managing their supply chains, with very effective results."

Jim Manton, president and COO of Pinacor, said: "The results that the [AMT] team delivered on the supply chain analysis helped Pinacor identify opportunities for optimizing the product flow between our companies. . . . I am pleased to see that both IBM and Pinacor are focusing on the recommendations to make the necessary improvements. . .”.

Mac McNeill, senior vice president of global operations for GE Capital IT Solutions, who sponsored a four-month project using the AMT to model GE Capital’s personal-computer supply chain commented: “The modeling allowed us to develop a base case using actual end-user customer sales and then to quickly model and optimize many alternatives based on various levels of GE forecast accuracy, IBM fill rates, transit times, in-bound and out-bound delays, and commonality of parts. The optimization results will allow us to develop action plans to balance improved levels of serviceability with lower levels of inventory.”