Multi-Echelon Inventory Optimization

By
Calvin B. Lee, Ph.D.
Vice President and Chief Scientist,
Evant Inc.

EVANT

Multi-Echelon Inventory Optimization

By Calvin B. Lee, Ph.D.
Vice President and Chief Scientist, Evant Inc.

The optimal deployment of inventory is a vital business function for an enterprise. The well-documented benefits of running a manufacturing, distribution or retailing operation with leaner inventory range from a permanent reduction in working capital to increased sales and higher customer satisfaction. As Forrester Research pointed out in a recent report, the ability to increase inventory turns is a key differentiator between highly successful and more poorly performing companies (e.g., Wal-Mart vs. Kmart; Dell vs. Compaq).

Managing inventory can be a daunting task for an enterprise with tens of thousands of products that are located in hundreds of locations. The challenge is even greater when the locations are situated in different tiers or echelons of the enterprise's distribution network. In such *multi-echelon networks*, new product shipments are first stored at a regional or central facility. These central facilities are the internal suppliers to the customer-facing locations. This is a common distribution model for many retail chains as well as for large distributors and manufacturers. For example, a large pharmaceutical wholesaler's distribution network consists of one regional distribution center (RDC) and more than 30 forward distribution centers (DCs). Another national retailer of automotive aftermarket parts and equipment manages more than 25 million Stock-Keeping Units (SKUs) that are spread out across 10 DCs and more than 900 stores. And finally, a global manufacturer/distributor of furniture fittings carries inventory in European DCs located near its factories before shipping the finished goods to 15 local DCs worldwide; these DCs serve end customers.

Managing inventory in a multi-echelon network vs. a single-echelon network presents major pitfalls. One is the failure to achieve true network inventory optimization, because replenishment strategies are applied to one echelon without regard to its impact on the other echelons. A network view of inventory usage up and down the demand chain is absent when you are only dealing with a single echelon of locations. Another pitfall is to base upper-echelon replenishment decisions on specious demand forecasts. These pitfalls can create substantial negative consequences, including the following:

- The network carries excess inventory in the form of redundant safety stock.
- End customer service failures occur even when adequate inventory exists in the network.
- Customer-facing locations experience undesirable stockouts, while service between echelons is more than acceptable.
- External suppliers deliver unreliable performance, because they have received unsatisfactory demand projections.
- Shortsighted internal allocation decisions are made for products with limited availability.

This paper will examine alternative approaches for tackling the thorny problem of managing inventory in a multi-echelon network and will present a method for minimizing inventory across echelons while simultaneously meeting all of your customer service goals.

Managing Inventory in Multi-Echelon Networks

The complexities of managing inventory increase significantly for a multi-echelon distribution network with multiple tiers of locations (e.g., a network comprising a central warehouse and downstream customer-facing locations). All locations are under the internal control of a single enterprise. Instead of simply replenishing the warehouse or the DCs that sit between your supplier and your end customers, as in the single-echelon situation, you also need to contend with the problems of replenishing another distribution point between your supplier and your DCs. For the purposes of this paper, we will refer to this additional distribution point as an RDC. The objective of multi-echelon inventory management is to deliver the desired end customer service levels at minimum network inventory, with the inventory divided among the various echelons.

Note: *We use the terms "RDC" and "DC" to distinguish the entities in the separate echelons. For a manufacturer or wholesaler, the terms "hub" and "spoke" could be equally applicable. For a retailer, the terms "central warehouse" and "store" might be relevant. For clarity, we will focus on the interaction between two echelons. The logic is easily extended to more than two echelons through recursion. In addition, we will use the term "enterprise" to refer to the distribution entity as a single, cohesive organization.*

Managing Inventory in Single-Echelon Networks

Before delving into the difficulties of managing the inventories in more than one echelon, let us review the single-echelon problem. In this environment, the distribution network is Supplier ➜ DC ➜ Customers. *Table 1* describes the inventory drivers for a SKU that is located at a DC.

In this single-echelon situation, the lead times are between the DC and its external supplier. The enterprise's order supply strategies depend on its internal cost factors—such as those associated with handling and carrying inventory—and the external supplier's ordering constraints and bracket discounts. For this reason, the replenishment quantities depend on a combination of internal and external factors.

INVENTORY	DESCRIPTION
Demand	Rate of product flow out of the DC
Demand Variation	Fluctuation of the product outflow from period to period
Lead Time	Expected time delay between ordering and having new product available to fulfill demand
Lead Time Variation	Fluctuation of the lead time from order to order
Replenishment Review Frequency	Frequency with which the DC checks its inventory position to see if a new order is needed
Order Supply Strategy	The DC's time supply objective, which depends on the economic trade-offs between carrying inventory, handling, transportation and purchase cost
Service Level Goal	The DC's service commitment to end customers
Inventory Position	The DC's available stock, taking into account the on-hand inventory, on-order quantities, back orders and committed stock

Table 1: Inventory Drivers for a SKU Located at a DC

Managing Inventory in Multi-Echelon Networks

Now consider the same product in a multi-echelon network that includes an RDC between the suppliers and the DCs. The same inventory drivers described in the preceding section apply for the SKU at the RDC location. However, some significant issues emerge:

- What is the proper measure of demand to the RDC, and how should this demand be forecasted?
- How do you measure the demand variation into the RDC?
- How does the trend toward larger orders from the RDC to the supplier affect the order supply strategy for the RDC SKU?
- What is the optimal service level goal between the RDC and its "customers," which are the DCs?
- How do you factor the individual DCs' inventory positions into the RDC replenishment decisions?
- How do the inventory drivers at the RDC, such as the replenishment review frequency and the service level goal, affect inventory and service levels at the DC level?
- When faced with a limited supply situation at the RDC, how should you allocate product down to the DCs?

Because the RDC also stocks inventory for the SKU, the replenishment decision at each DC also must address some *new* questions because of its relationship with an internal supplier:

- How will the ordering constraints imposed by the RDC (not the supplier) influence the DC's order supply strategy?
- How will different DC replenishment review frequencies and alternative order supply strategies affect the RDC? How do you factor the RDC service level goal into the DC replenishment strategy?
- To achieve the targeted service level commitment with its end customers, should the DC use the same service level goal when the RDC is available as a backup source for end customers?
- How can you use the RDC in an expedited ordering process?
- Do the external supplier lead time and lead time variation still play a role in the DC's replenishment strategy?

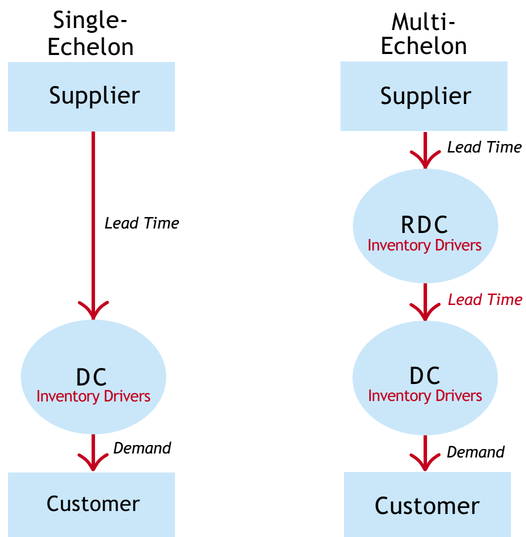

Figure 1: Inventory Drivers

Figure 1 illustrates how the inventory drivers are linked in the two echelons. Circular nodes represent locations that are under the control of the one enterprise. The node labeled "DC" stands for *all* the DCs that stock the same product. The inventory drivers, denoted in red, are controllable by the enterprise. That is, the replenishment review frequency, the order supply strategy and the service level goal are replenishment control variables that a decision maker can set to influence the amounts of inventory to be carried and the service levels offered to downstream customers. The approaches for setting these control variables in the single-echelon case are well known, but how should you set the control variables in the multi-echelon case? Today, enterprises typically use one of two approaches:

1. Apply the single-echelon approach to each echelon in the network.
2. Use a distribution requirements planning (DRP) approach or a variation.

Using a Sequential Approach

The first method follows the obvious route. You take the established approach used in the single-echelon case and apply it twice—once to replenish the DCs based on their inventory drivers, and then again to replenish the RDC, based on its inventory drivers. In this case, the DCs would use the end customer demands and the RDC ➜ DC lead times. The lead times for the RDC are clear; they are simply the lead times to the supplier. But how should demand forecasts be created for the RDC? One way is to base demand on historical orders from the DCs to the RDC. Another way is to simply "pass up" the end customer demands from DCs. Both ways are flawed, as we shall see.

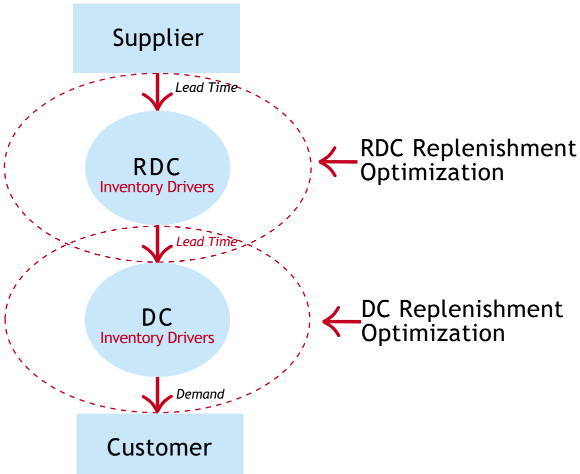

Figure 2: Sequential Approach

Figure 2 illustrates the DC and RDC replenishment approach that was just described. This is a *sequential approach* that splits multi-echelon replenishment into two separate challenges—one for the DC and one for the RDC. This approach presents many problems:

- **Lack of visibility *up* the demand chain.** When a DC replenishes itself, it is oblivious to suppliers beyond the RDC. In particular, the DC ignores any lead times except the lead time from the RDC. The DC also assumes that the RDC will fill its replenishment orders completely every time. Finally, the DC has no visibility into the RDC's inventory balance.
- **Lack of visibility *down* the demand chain.** When the RDC replenishes itself, it is oblivious to customers beyond those of the DCs. In addition, the RDC has no visibility into the DCs' inventory balances or demands.
- **Demand distortion from the bullwhip effect.** Because the RDC and the DCs develop independent demand forecasts (based on their immediate customers' demands), the *bullwhip effect* causes increased demand variation between the RDC and the DCs. This results in unnecessary inventory at the RDC.
- **Total network costs remain unevaluated.** When the RDC's or the DCs' replenishment strategy changes (e.g., by altering one of the controllable inventory drivers), the cost implications of the new strategy on the other echelon are not considered. The focus is strictly on the impact on the echelon at hand.

Distribution Requirements Planning

The distribution requirements planning (DRP) approach is an extension of the materials requirements planning (MRP) method that manufacturers rely on to determine and satisfy requirements for the components used to assemble complex products. The parts or subassemblies used to make the finished product are dependent demand items. Instead of managing dependent demands in different stages of an assembly operation, as MRP does, DRP manages demands for a product that is located at different echelons of the distribution network. Demands for the product in a higher echelon are, in a sense, dependent on lower-echelon demands for the same product. With a DRP approach, DC-level demand forecasts are first used to develop gross product requirements. These forecasts are combined with safety stock requirements and stock status information to arrive at net requirements at the DCs. This is analogous to an MRP master schedule. The time-phased dependent demand at the RDC is calculated by offsetting the DC net requirements by the RDC ➜ DC lead times and summing over corresponding time periods. The RDC uses these "pass-up" demands to replenish itself.

The DRP approach has several major shortcomings. The No. 1 weakness is the deterministic perspective vis-à-vis pass-up demands and lead times. An immediate consequence of this is the subjective way in which the RDC safety stock is usually determined. Because the requirements passed up to the RDC include no uncertainty, there is no rigorous method for determining safety stock. This is why enterprises that use this replenishment method generally use rules of thumb for the RDC safety stock; this unscientific approach leads to excess inventory. It is not surprising that safety stock determination is somewhat loose—DRP has its roots in manufacturing, where production and transportation costs are of greater concern than inventory costs. Like the sequential approach, DRP fails to exploit visibility up the demand chain and lacks a network view of inventory optimization.

In particular, there is no linkage between the safety stocks in the two echelons. Therefore, any attempt to optimally balance inventory between the two echelons is impracticable.

Assessing the Sequential and DRP Approaches

Both the sequential and the DRP approach result in excess inventory without necessarily improving service levels to the end customers. Although each echelon might be able to achieve reasonable results—from its own myopic viewpoint of the total problem—the results are not necessarily the optimal solution for the total network. That is, the total inventory of the RDC and at the DCs is not minimized in the pursuit of end customer service objectives.

Bullwhip Effect in Multi-Echelon Networks

Much has been written about the bullwhip effect and the way it distorts demand information as the information is transmitted up a demand chain. Although the bullwhip effect can appear in single-echelon situations, it is usually outside the enterprise's control. However, this is not true in multi-echelon networks; in these cases, the enterprise must consider and manage the bullwhip effect. The bullwhip effect is caused by independent rational decisions in demand signal processing, order batching, reactions to price variations and shortage gaming. A sequential approach falls into the bullwhip trap by using multiple independent demand forecasts in different echelons. Order batching in a lower echelon causes extra demand variability between echelons. Finally, the lack of visibility up and down the demand chain in a sequential or DRP approach can produce inventory stockpiling to achieve customer service objectives. A multi-echelon network offers the opportunity to properly measure the bullwhip effect, to identify its root causes and to reduce or eliminate its impact on demand chain performance. To ignore this opportunity is to let the bullwhip effect degrade forecast accuracy, increase inventory, increase operational costs and lower customer service levels.

The True Multi-Echelon Approach

When an enterprise with multiple tiers of locations uses a true multi-echelon approach to manage inventory, the primary objective is to minimize the total inventory in all echelons (the RDC and all the DCs) while meeting service commitments to end customers. Even though inventory is the main focus, transportation and warehouse operations expenses also are kept in line, because their cost factors are part of the optimization. Both the sequential and the DRP approaches treat each echelon as a separate problem without considering the inventory impact one echelon may exert on another. With a multi-echelon approach, demand forecasting and inventory replenishment decisions are made at the enterprise level in a single optimization exercise rather than in a sequence of sub-exercises for each echelon. Specifically, a true multi-echelon approach should:

- **Avoid multiple independent forecast updates in each echelon.** The primary customer demand signal and other information at the DCs drive the forecasts in all echelons. A true multi-echelon approach eliminates the reliance on demands from the immediate downstream customer.
- **Account for all lead times and lead time variations.** In each echelon, the replenishment decisions account for lead times and lead time variations of all upstream suppliers, not just the immediate suppliers.
- **Monitor and manage the bullwhip effect.** The enterprise measures the demand distortion and determines the root causes for possible corrective actions.

- **Enable visibility *up* and *down* the demand chain.** Each echelon takes advantage of visibility into the other echelon's inventory positions—what is on hand, on order, committed and back ordered. At the DCs, this negates any need for shortage gaming. At the RDC, visibility into DC inventories improves projections of demand requirements.
- **Synchronize order strategies.** Synchronizing the ordering cycles at the DCs with RDC operations reduces lead times and lead time variation between the RDC and the DCs. Multi-echelon models can evaluate the impact on both echelons of different synchronization strategies.
- **Offer differentiated service levels.** The RDC can provide different service levels (for the same product) to different DCs. A multi-echelon approach makes this possible, because the enterprise controls how and when a product *enters* and *leaves* the RDC.
- **Correctly model the inter-active effects of alternative replenishment strategies of one echelon on another.** Alternative strategies include different replenishment review frequencies, order supply strategies, service level goals and SKU stratifications.

Figure 3 illustrates how *all* of the inventory drivers in both echelons are factored into the DC and RDC replenishment decisions to truly optimize the network's inventory.

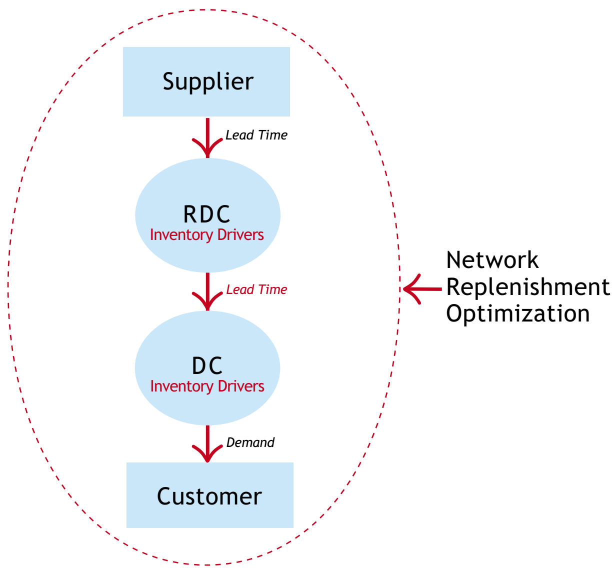

Figure 3: Multi-Echelon Approach

Table 2 summarizes the key differences between the multi-echelon solution and other approaches.

"NWDCo" Case Study: Background

Until a few years ago, the network of NWDCo, a national wholesale distributor, comprised only DCs situated between NWDCo's suppliers and customers. Each DC would order product from the suppliers and fill orders to customers. The DCs are customer-facing

KEY AREAS	SEQUENTIAL APPROACH	DRP APPROACH	TRUE MULTI-ECHELON APPROACH
Optimization Objective	Meet immediate customer's (RDC or DC) service goals at minimum inventory; suboptimal for network	Not optimization; objective is to provide net requirements upstream to determine replenishment needs	Meet end-customer service goals at minimum inventory
Demand Forecasting	Independent forecasts in each echelon based on immediate customer's demands	Pass-up demands or projected orders with no measures of their variabilities	Forecasts based on lowest echelon's primary demand signals and other information; demand variations also are forecasted
Lead Times	Uses immediate suppliers' lead times and lead time variabilities	Uses immediate suppliers' lead times; ignores variabilities	Uses all lead times and lead time variations of upstream suppliers
Bullwhip Effects	Ignored	Ignored	Effects measured and accounted for in overall replenishment strategy
Network Visibility	*Immediate* downstream customers' demands and *immediate* upstream suppliers' lead times—*myopic* view of the network	Some downstream visibility; no upstream visibility	All echelons have complete visibility into other echelons; this visibility is exploited in the replenishment logic
Order Synchronization Between Echelons	Ignored	Maybe, probably not	Fully modeled to reduce unnecessary lags in network
Differentiated Customer Service	Not possible	Not possible	Achievable, as orders out of a higher echelon location to a lower echelon are fully controllable; allocation schemes using set-aside inventories can be used
Cost Implications Between Echelons	Not possible	Not possible	Fully modeled so true network optimization can be achieved

Table 2: True Multi-Echelon vs. Other Approaches

and have average service commitments of more than 98%. A DC typically stocks between 25,000 and 50,000 SKUs. Today, NWDCo has several RDCs and the same DCs. Each DC is assigned to one of the RDCs. Every item that is stocked at a DC is either a *vendor-direct* or an *RDC item*. A vendor-direct item is replenished directly from an external supplier; it does not flow through any RDC. An RDC item at a DC is internally replenished from an RDC. Roughly half the company's sales are tied to items that are stocked at both the RDC and DC echelons. An RDC stocks about 10,000 SKUs.

Each RDC has its own customers that order directly from the RDC. These customers require 99% service levels. Thus, each RDC has two types of customers—external customers and internal customers represented by the DCs. Each DC orders from its RDC on a fixed schedule, but the external customers order any time they want.

The lead times from the suppliers to the RDCs are generally shorter and less erratic than the lead times from the suppliers to the DCs had been before the RDCs existed. Instead of ranging from one to three weeks, the lead times are now three to 10 days. The time-definite deliveries between the RDCs and the DCs range from one to three days, depending on the DC.

Challenges

NWDCo's end customers have very high service requirements. One way NWDCo can assure that it meets these high service goals is to carry extra inventory at the DCs. At the RDCs, inventories are carried to replenish the DCs and to satisfy the RDCs' external customers. The problem is to deploy the optimal balance of inventory in the two echelons. NWDCo can use many control levers to shift the inventory balance between its RDCs and DCs. How should NWDCo set these to achieve service objectives at minimum inventory? Can slow-moving products simply be centralized at the RDCs with little or no product at the DCs?

The inventory management problem is further complicated because of the high service requirements of the RDCs' external customers. The RDCs must simultaneously set aside enough inventory to satisfy external customers' demands *and* to provide the DCs with enough product on hand to satisfy the DCs' customers' demands.

An overriding question NWDCo must address is how to come up with the right demand forecasts at the RDCs. These forecasts are crucial to replenishment decisions at the

RDCs. Should the DC demands simply be passed up to the RDCs for use in determining statistical forecasts? Should the RDCs simply aggregate demand forecasts that are determined at the DCs? Should the RDCs develop statistical forecasts based on historical shipments to the DCs? What about historical orders from the DCs?

Other challenges exist. NWDCo needs to attain and fully exploit information transparency in its network. This means the RDCs and the DCs must freely share information on demand, inventory, in-transit orders and factors that impact demand, such as promotions activity. Another problem is to synchronize the replenishment activities between the two echelons to minimize lead times and to balance orders between DCs and RDCs over the days of the week. Failure to adequately tackle these challenges will lead to increased inventory and potential service failures because of the bullwhip effect.

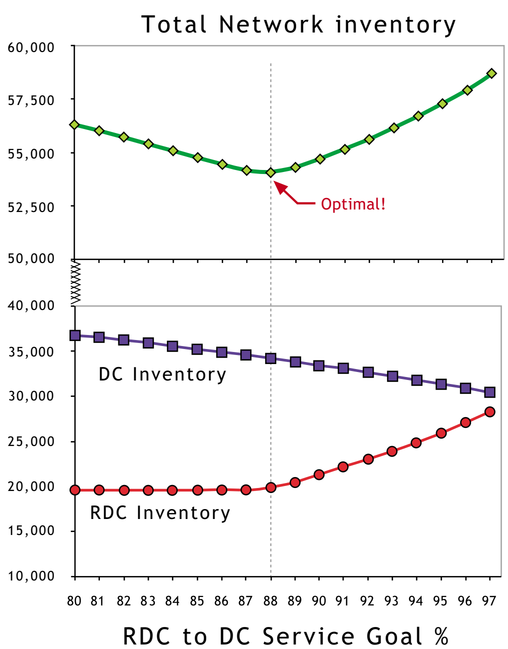

Figure 4: Finding the Optimal RDC-to-DC Service Level Goal

Solution

NWDCo implemented a true multi-echelon inventory management system in phases, with each phase incorporating additional vendor lines. All replenishment decisions now exploit network visibility. The system models all cost implications of replenishment strategies at the DCs on the RDCs and vice versa. There are direct linkages between inventory drivers in one echelon and inventory levels in the other echelon. As an illustration, *Figure 4* depicts how NWDCo sets one of the control variables, the *RDC ➜ DC service level goal*, to minimize the total inventory at the RDC and DCs. As expected, the RDC inventory increases as this parameter increases. Less obvious is the effect on DC inventories. As the expected fill rate from the RDC becomes larger, the total inventory at the DCs decreases. The optimal setting for this parameter is the value that minimizes the total inventory in both echelons. Obtaining this optimal setting requires careful modeling of the interactive effects on both the RDC and DC safety stocks. NWDCo follows this optimization process for every RDC item.

Benefits

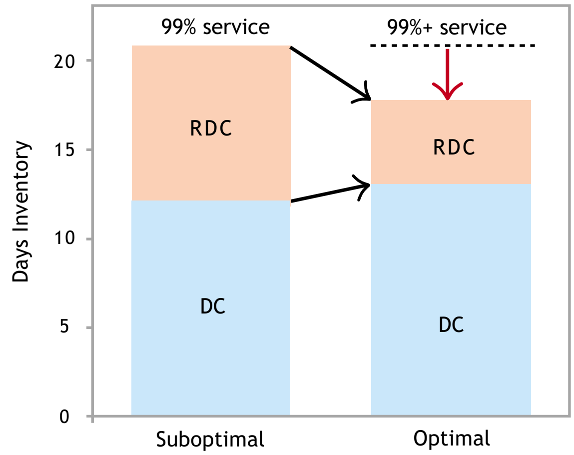

Figure 5: Driving Down Network Inventory with Multi-Echelon Optimization

By using a true multi-echelon approach, NWDCo makes optimal replenishment decisions for all SKUs across every echelon. This enables the enterprise to attain or exceed service goals while driving down inventory requirements in the total network. *Figure 5* illustrates a typical result for a class of SKUs with average velocity. Inventory is properly balanced between the two echelons to reduce network inventory by more than three days while maintaining or exceeding 99% customer service goals. Reduced inventories free up working capital for the enterprise. Besides increasing cash flow, the increased cash helps to permanently raise earnings for NWDCo. At the same time, higher service levels generate increased revenue and customer satisfaction.

Conclusion

A multi-echelon distribution network presents many opportunities for inventory optimization that the enterprise must pursue to offset potential increases in transportation, warehouse and occupancy costs. The key to achieving those savings is to use a true multi-echelon strategy to manage inventory. It is not a simple task to pursue such a strategy because of the multiplicities of inventory drivers and the complexities in modeling the interactions of the drivers between echelons. Nevertheless, the benefits are worth the effort. Taking the right approach can yield rewards on both sides of the inventory equation—better customer service with less inventory. Using a true multi-echelon approach is the ultimate win-win strategy for inventory management.

Author Bio

Calvin B. Lee has more than 25 years of experience in the transportation, wholesale distribution and retail industries. Prior to joining Evant Inc., Dr. Lee was Director of Merchandising Logistics at Webvan Group, Inc. Previously, he was Vice President of Operations Research at McKesson Corp., where he made major contributions to inventory management and distribution network redesign initiatives. Dr. Lee managed the analytic services team at Southern Pacific Transportation (now part of Union Pacific) and oversaw its operations research and decision support activities. Currently, he is on the adjunct faculty of several San Francisco Bay Area universities. Dr. Lee holds a Ph.D. in Applied Mathematics and an M.A. in Mathematics, both from the University of California, Berkeley.