Chapter 1

Warehouse rationale

1.1 Why have a warehouse?

Why have a warehouse at all? A warehouse requires labor, capital (land and storage-and-handling equipment) and information systems, all of which are expensive. Is there some way to avoid the expense? For most operations the answer is no. Warehouses, or their various cousins, provide useful services that are unlikely to vanish under the current economic scene. Here are some of their uses:

To better match supply with customer demand: One of the major challenges in managing a supply chain is that demand can change quickly, but supply takes longer to change. Surges in demand, such as seasonalities strain the capacity of a supply chain. Retail stores in particular face seasonalities that are so severe that it would be impossible to respond without having stockpiled product. For example, Toys R Us does, by far, most of its business in November and December. During this time, their warehouses ship product at a prodigious rate (some conveyors within their warehouses move at up to 35 miles per hour). After the selling season their warehouses spend most of their time building inventory again for the following year. Similarly, warehouses can buffer the supply chain against collapsing demand by providing space in which to slow or hold inventory back from the market.

In both cases, warehouses allow us to respond quickly when demand changes. Response-time may also be a problem when transportation is unreliable. In many parts of the world, the transportation infrastructure is relatively undeveloped or congested. Imagine, for example, shipping sub-assemblies to a factory in Ulan Bator, in the interior of Asia. That product must be unloaded at a busy port, pass through customs, and then travel by rail, and subsequently by truck. At each stage the schedule may be delayed by congestion, bureaucracy, weather, road conditions, and so on. The result is that lead time is long and variable. If product could be warehoused in Shanghai or closer to the point of use, it could be shipped more quickly, with less variance in lead time, and so provide better customer service.

Warehouses can also buffer against sudden changes in supply. Vendors may give a price break to bulk purchases and the savings may offset the expense of storing the product. Similarly, the economics of manufacturing may dictate large batch sizes to amortize large setup costs, so that excess product must be stored. Similarly, warehouses provide a place to store a buffer against unreliable demand or price increases.

To consolidate product to reduce transportation costs and to provide customer service.

There is a fixed cost any time product is transported. This is especially high when the carrier is ship or plane or train; and to amortize this fixed cost it is necessary to fill the carrier to capacity. Consequently, a distributor may consolidate shipments from vendors into large shipments for downstream customers. Similarly, when shipments are consolidated, then it is easier to receive downstream. Trucks can be scheduled into a limited number of dock doors and so drivers do not have to wait. The results are savings for everyone.

Consider, for example, Home Depot, where more than a thousand stores are supplied by several thousands of vendors. Because shipments are frequent, no one vendor ships very much volume to any one store. If shipments were sent direct, each vendor would have to send hundreds of trailers, each one mostly empty; or else the freight would have to travel by less-than-truckload carrier, which is relatively expensive. But there is enough volume leaving each vendor to fill trailers to an intermediate crossdock. And each crossdock receives product from many vendors for each store, so that the the total freight bound for each store is typically sufficient to fill a trailer. The result is that vendors send fewer shipments and stores receive fewer shipments. Moreover, the freight will have traveled by full-truck-load and so paid significantly less transportation costs.

Another example of the benefits of consolidation is when the differentiation of product is postponed, typically within the warehouse. For example, the finished-goods warehouse of a manufacturer of pet food held 1,500 different products, which represented only 25 different recipes. By holding the product as "bright stack" (unlabeled cans), the product differentiation was postponed until the customer placed an order, when the appropriate cans would be labeled and shipped. This postponement allowed the safety stocks of the 1,500 products to be consolidated into only 25 safety stocks, which were easier to manage and could be reduced due to pooling of risk.

To provide value-added processing: Increasingly, warehouses are being forced to assume value-added processing such as light assembly. This is a result of manufacturing firms adopting a policy of postponement of product differentiation, in which the final product is configured as close to the customer as possible. Manufacturers of personal computers have become especially adept at this. Generic parts, such as keyboards, disk drives, and so on, are shipped to a common destination and assembled on the way, as they pass through a warehouse or the sortation center of a package carrier. This enables the manufacturer to satisfy many types of customer demand from a limited set of generic items, which therefore experience a greater aggregate demand, which can be forecast more accurately. Consequently safety stocks can be lower. In addition, overall inventory levels are lower because each item moves faster.

Another example is pricing and labeling. The state of New York requires that all drug stores label each individual item with a price. It is more economical to do this in a few warehouses, where the product must be handled anyway, than in a thousand retail stores, where this could distract the workers from serving the customer.

1.2 Types of warehouses

Warehouses may be categorized by type, which is primarily defined by the customers they serve. Here are some of the more important distinctions:

A *retail distribution center* typically supplies product to retail stores, such as Wal-Mart or Target. The immediate customer of the distribution center is a retail store, which is likely to be a regular or even captive customer, receiving shipments on regularly scheduled days. A typical order might comprise hundreds of items; and because the distribution center might serve hundreds of stores, the flow of product is huge. The suite of products changes with customer tastes and marketing plans.

A *service parts distribution center* is among the most challenging of facilities to manage. They typically hold spare parts for expensive capital equipment, such as automobiles, airplanes, computer systems, or medical equipment. Consequently, a typical facility contains a huge investment in inventory: tens of thousands of parts, some very expensive. (A typical automobile contains almost 10,000 parts.) Because of the large number of parts, total activity in the DC may be statistically predictable, but the demand for any particular part is relatively small and therefore hard to predict. This means that the variance in requests is large and so large quantities of safety stock must be held. Furthermore, when a part is requested, it is generally very urgent, because an important piece of capital equipment might be unusable, such as a truck or a MRI device. To further compound the challenges, there is typically long lead times to replenish parts to the warehouse.

It is not unusual for customer orders to fall into one of two distinct categories: Some customers, such as vehicle dealers or parts resellers, typically submit large orders, for 10's or 100's of different skus; while other customers, such as independent repair shops, might order only what they need to repair a single vehicle. Worse, customers ordering for repair might order before they are absolutely sure which parts need replacement; and so there can be a significant percentage of returns to be handled at the warehouse.

A *catalog fulfillment* or *e-commerce* distribution center typically receives small orders from individuals by phone, fax, or the Internet. Orders as typically small, for only 1–3 items, but there may be many such orders, and they are to be filled and shipped immediately after receipt.

A *3PL warehouse* is one to which a company might outsource its warehousing operations. The 3PL provider might service multiple customers from one facility, thereby gaining economies of scale or complementary seasons that the customers would be unable to achieve on their own. 3PL facilities may also be contracted as overflow facilities to handle surges in product flow.

While there are many types of warehouses in the supply chain, one of the main themes of this book is that there is a systematic way to think about a warehouse system regardless of the industry in which it operates. As we shall show, the selection of equipment and the organization of material flow are largely determined by

- Inventory characteristics, such as the number of products, their sizes, and turn rates;
- Throughput and service requirements, including the number of lines and orders shipped per day;
- The footprint of the building and capital cost of equipment.

Chapter 2

Material flow

Here we briefly discuss a few issues that help lay the foundations for warehouse analysis.

2.1 The fluid model of product flow

The “supply chain” is the sequence of processes through which product moves from its origin toward the customer. In our metaphor of fluid flow we may say that warehouses represent storage tanks along the pipeline.

The analogy with fluid flows can also convey more substantial insight. For example, consider a set of pipe segments of different diameters that have been joined in one long run. We know from elementary fluid dynamics that an incompressible fluid will flow faster in the narrower segments of pipe than in the wider segments. This has meaning for the flow of product: The wider segments of pipe may be imagined to be parts of the supply chain with large amounts of inventory. On average then, an item will move more slowly through the region with large inventory than it will through a region with little inventory.

The fluid model immediately suggests other general guidelines to warehouse design and operation, such as:

- Keep the product moving; avoid starts and stops, which mean extra handling and additional space requirements.
- Avoid layouts that impede smooth flow.
- Identify and resolve bottlenecks to flow.

Later we shall rely on the fluid model to reveal more profound insights.

It is worth remarking that the movement to “just-in-time” logistics is roughly equivalent to reducing the diameter of the pipe, which means product flows more quickly and so flow time and in-transit inventory are reduced (Figure 2.1).

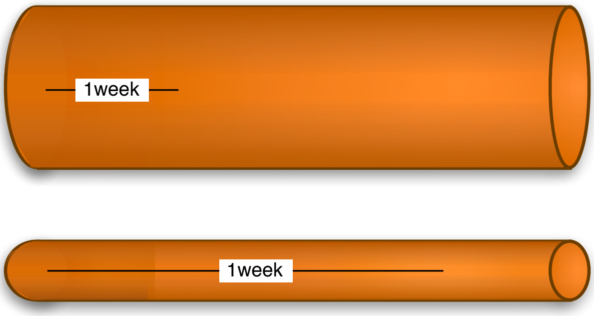

Figure 2.1: If two pipes have the same rates of flow, the narrower pipe holds less fluid. In the same way, faster flow of inventory means less inventory in the pipeline and so reduced inventory costs.

2.2 Units of handling

Even though it is a frequently useful metaphor, most products do not, of course, flow like incompressible fluids. Instead, they flow more like a slurry of sand and gravel, rocks and boulders. In other words, the product is not infinitely divisible but rather is granular at different scales.

A "stock keeping unit" is the smallest physical unit of a product that is tracked by an organization. For example, this might be a box of 100 Gem Clip brand paper clips. In this case the final customer will use a still smaller unit (individual paper clips), but the supply chain never handles the product at that tiny scale.

Upstream in the supply chain, product generally flows in larger units, such as pallets; and is successively broken down into smaller units as it moves downstream, as suggested in Figure 2.2. Thus a product might move out of the factory and to regional distribution centers in pallet-loads; and then to local warehouses in cases; and finally to retail stores in inner-packs or even individual pieces, which are the smallest units offered to the consumer. This means that our fluid model will be most accurate downstream, where smaller units are moved.

2.3 Storage: "Dedicated" versus "Shared"

Each storage location in a warehouse is assigned a unique address. This includes both fixed storage locations, such as a portion of a shelf and mobile locations such as the forks of a lift truck. Storage locations are expensive because they represent space, with consequent costs of rent, heating and/or air-conditioning, security, and so on. In addition, storage locations are typically within specialized

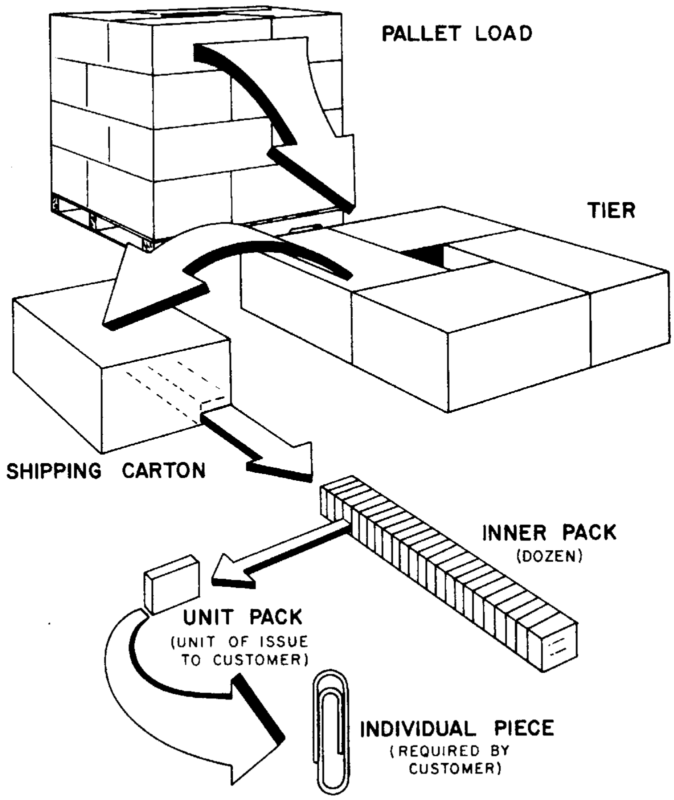

Figure 2.2: A product is generally handled in smaller units as it moves down the supply chain. (Adapted from "Warehouse Modernization and Layout Planning Guide", Department of the Navy, Naval Supply Systems Command, NAVSUP Publication 529, March 1985, p 8–17).

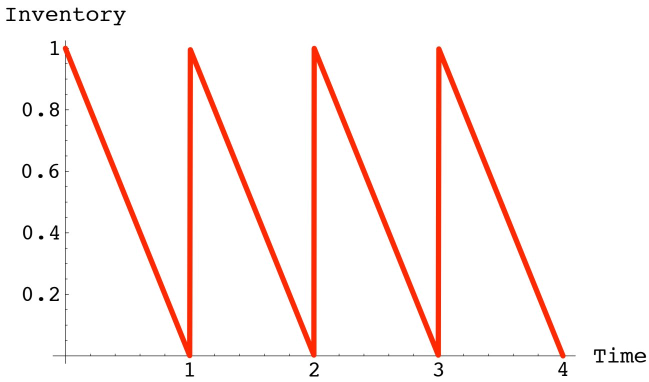

Figure 2.3: An idealization of how the inventory level at a location changes over time

equipment, such as shelving or flow rack, which are a capital cost. These costs impel us to use storage space as efficiently as possible.

There are two main strategies used in storing product. The simplest is *dedicated* storage, in which each location is reserved for an assigned product and only that product may be stored there. Because the locations of products do not change, more popular items can be stored in more convenient locations and workers can learn the layout, all of which makes order-picking more efficient.

The problem with dedicated storage is that it does not use space efficiently. This can be seen by tracking the amount of inventory in a given location. If we plot the inventory level, measured for example by volume, we would see a sawtooth shape such as in Figure 2.3 (which represents an idealization of the inventory process.) In one cycle, the storage location is initially filled but empties as product is withdrawn to send to customers. As a result, on average this storage location is half empty.

A warehouse may have thousands or tens-of-thousands of storage locations. If using dedicated storage, each will have an assigned product. Each product may have a different replenishment cycle and so, upon entering such a warehouse, one expects to see many storage locations that are nearly empty, many that are half-full, and many that are nearly full. On average the storage capacity is only about 50% utilized.

To improve on this, one can use a strategy of *shared* storage. The idea here is to assign a product to more than one storage location. When one location becomes empty, it is available for reassignment, perhaps to a different product. This space then can be filled again, rather than waiting until the original product is replenished (presumably when the last of the warehouse supply has been exhausted). The more storage locations over which a product is distributed, the less product in each location, and so the sooner one of those locations is emptied and the sooner that space is recycled. Therefore we expect better utilization of space when shared storage is used.

Unfortunately, shared storage also has some disadvantages. Most immediately, the locations of products will change over time as locations are emptied and restocked with other products. This means that workers cannot learn locations and so must be directed to locations by a warehouse management (software) system. Another disadvantage is that it becomes more time-consuming to put away newly received product because it has to be taken to more locations. There can be other, social complications as well. For example, imagine an order picker who has been directed to the other side of the warehouse to pull a product for a customer. That order picker may be tempted to pick the product from a more convenient location, thus creating discrepancies between book and physical inventory at two locations. For these reasons, shared storage requires greater software support and also more disciplined warehouse processes.

Shared storage is generally more complicated to manage because it introduces many possible trade-offs. In particular, one can manage the trade-off between space and time (labor) on an activity-by-activity basis. For example, one can retrieve product from the least-filled location (to empty and recycle that location as soon as possible) or from the most convenient location (to save labor). Similarly, one can replenish product to the most nearly empty location to fill that empty space or to the most convenient location to save labor time.

How much improvement in space utilization is possible with shared storage? Consider a product that is requested at a constant rate, as in our idealization of Figure 2.3. Suppose we hold two weeks supply of this product. If we store it in two locations of equal size and direct all order-picking to only one location then after one week, the first location will be emptied and available for reassignment. After the second week the second location will be emptied and available for reassignment. During the first week, the first location was half full on average and the second location was completely full. During the second week the second location was half-full on average. Thus, on average, we used two of the three location-weeks of storage space assigned to this product for an average utilization of 66%, which is better than the 50% estimated for dedicated storage. More improvement is possible if a product is stored in more locations, though the improvement diminishes and, moreover, the practical problems of management increase.

More generally we can argue as follows. The average space utilization is the average inventory divided by the average space required. Assume for convenience that demand is constant and the sku has been stored in k locations of identical size. Then the inventory cycle may be imagined as consisting of k periods, as in Figure 2.4. The average space utilization is then

$$\begin{aligned}\frac{k/2}{(k+(k-1)+\cdots+2+1)/k} &= \frac{k/2}{(k\,(k+1)\,/2)\,/k} \\ &= \frac{k/2}{(k+1)\,/2}\end{aligned}$$

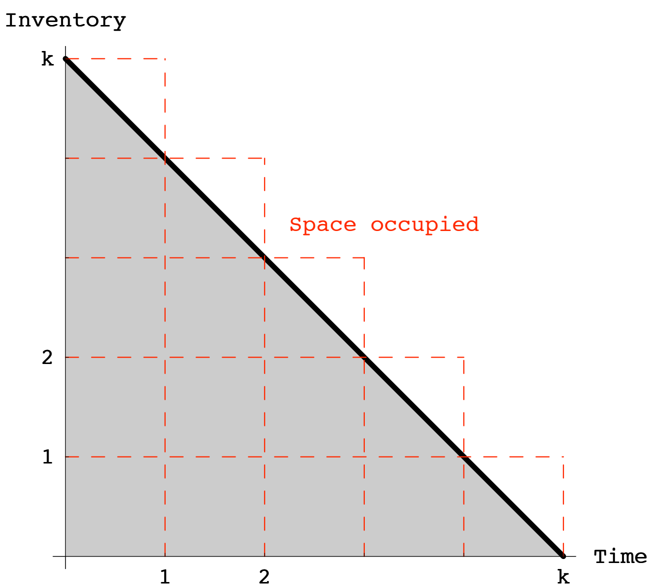

Figure 2.4: Use of k locations to hold product under a policy of shared storage

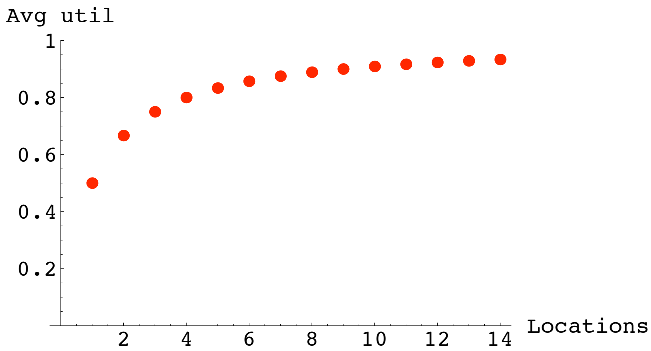

Figure 2.5: Space utilization increases with additional storage locations under shared storage, but at a diminishing rate.

$$= \frac{k}{k+1}$$

For this simple model then, in which demand is constant, and all picks are from a single active pick location until that location is exhausted:

Theorem 2.1 *When a sku is stored in k locations of equal size the average space utilization is $k/k+1$.*

Figure 2.5 shows how the space utilization of storing a sku in more locations increases but with diminishing returns.

Interestingly, there is a slight sampling bias if one were to examine space utilization directly by taking a census within the warehouse. This is the subject of Exercise 2.5 and it suggests that you will tend to measure slightly smaller values than predicted by Theorem 2.1.

In practice, a strategy of shared storage is typically used in the bulk storage areas, where most of the volume of product is held on pallets. Dedicated storage may be used in the most active picking areas, which are much smaller than the bulk storage. Thus one gets efficient use of most of the space (bulk storage) with labor benefits where it matters most (active picking areas).

There are also hybrid schemes in which regions are reserved for groups of skus, but locations are not reserved. For example, an aisle might be reserved for skus of one type or from one vendor, but within that aisle, space would be shared amongst the skus.

2.4 The warehouse as a queuing system

A *queuing system* is a model of the following structure: Customers arrive and join a queue to await service by any of several servers. After receiving service the customers depart the system.

A fundamental result of queuing theory is known as Little's Law, after the man who provided the first formal proof of a well-known piece of folk-wisdom [28].

Theorem 2.2 (Little's Law) *For a queuing system in steady state the average length L of the queue equals the average arrival rate λ times the average waiting time W. More succinctly:*

$$L = \lambda W.$$

A warehouse may be roughly modeled as a queuing system in which skus are the customers that arrive at the receiving dock, where they join a queue (that is, are stored in the warehouse) to wait for service (shipping). If the warehouse is at steady state the product will be shipped at the same average rate at which it arrives. Then Little's Law applies and the average amount of product in the warehouse equals the arrival rate of product multiplied by the average time product is resident in our warehouse.

Here is an example of how we can use Little's Law to tease out information that might not be immediately apparent. Consider a warehouse with about 10,000 pallets in residence and that turn an average of about 4 times a year. What labor force is necessary to support this? By Little's Law:

$$10{,}000 \text{ pallets} = \lambda(1/4 \text{ year}).$$

so that

$$\lambda \approx 40{,}000 \text{ pallets/year}.$$

Assuming one 8-hour shift per day and about 250 working days per year, there are about 2,000 working hours per year, which means that

$$\lambda \approx 20 \text{ pallets/hour}.$$

Notice what we have just done: From a simple count of pallets together with an estimate of the number of inventory turns per year we estimated the labor requirements.

Little's Law can be very useful like this. Another typical use would be to compute an estimate of inventory turns after simply walking through the distribution center: One can estimate inventory (queue length) by counting storage positions, and rate of shipping (throughput) by observing the shipping dock, and then apply Little's Law.

What makes Little's Law so useful is that it continues to hold even when there are many types of customers, with each type characterized by its own arrival rate λ_i, waiting time W_i, and queue length L_i. Therefore the law may be applied to a single sku, to a family of skus, to an area within a warehouse, or to an entire warehouse.

One way of understanding Little's Law is as a simple identity of accounting. Divide a fixed period of time into n equally spaced intervals. Let $A(n)$ denote the total number of arrivals during this period of time, and let T_j denote the time in the system of the j^{th} arrival (assumed to be an integer number of periods). Arrivals occur only at the beginning of a period. Let I_i denote the inventory in the system at time i. Assume for now that $I_0 = I_n = 0$. If each arrival (customer) must pay 1 dollar per period as rent at the end of each period of stay, how much money does the system collect? On the one hand customer j pays T_j dollars, and so the answer is $\sum_{j=1}^{A(n)} T_j$. On the other hand, the system collects I_i dollars each period, and so the answer must also be $\sum_{i=1}^{n} I_i$. Therefore,

$$\sum_{i=1}^{n} I_i = \sum_{j=1}^{A(n)} T_j,$$

or, equivalently,

$$\frac{\sum_{i=1}^{n} I_i}{n} = \left(\frac{A(n)}{n}\right)\left(\frac{\sum_{j}^{A(n)} T_j}{A(n)}\right).$$

This may be interpreted as an approximation to Little's Law with queue length $L_n = \sum_{i=1}^{n} I_i/n$, arrival rate $\lambda_n = A(n)/n$ and average time in the system $W_n = \sum_{j=1}^{A(n)} T_j/A(n)$. Assuming the system is in steady state, as $n \to \infty$, $\lambda_n \to \lambda$, $W_n \to W$, and $L_n \to \lambda W$.

2.5 Questions

Question 2.1 *What are the five typical physical units-of-measure in which product is handled in a warehouse? For each unit-of-measure, state whether there are any standardized dimensions and, if so, identify them.*

Question 2.2 *In what ways has the inventory process depicted in Figure 2.3 been idealized?*

Question 2.3 *Consider a sku that has been allocated more than one location within a warehouse that is organized by "shared storage".*

- *What are the advantages/disadvantages of allowing picking from any or all of the locations?*
- *What are the advantages/disadvantages of restricting picking to only one of those locations (until empty, when another location would be designated the one from which to pick)?*

Question 2.4 *In real life a certain amount of safety stock may be held in a storage location to guard against stockout while awaiting replenishment to that location (which would interrupt order-picking). How would this affect the average utilization of storage space?*

Question 2.5 (Harder) *What value of space utilization would you expect to observe if you showed up at random to examine storage at a warehouse using shared storage? For simplicity, consider a single sku that is stored in k locations of identical size. Assume an idealized inventory process such as depicted in Figure 2.3 and assume that all picks are directed to the location that is most nearly empty.*

Prove that the expected space utilization under the model above (random sampling) is never greater than the mean space utilization.

Where is the difference in the mean and expected mean values greatest?

Explain how this difference arises.

Question 2.6 *Why does the model of Question 2.5 break down (that is, lose practical meaning) for very large numbers of storage locations?*

Question 2.7 *Your third-party warehouse has space available for 10,000 pallets and you have 20 forklift operators per 8-hour day for 250 working days a year. If the average trip from receiving to storage to shipping is 10 minutes, how many inventory turns a year could you support for a full warehouse?*

Question 2.8 *Your third-party warehouse is bidding for a contract to store widgets as they are manufactured. However, widgets are perishable and should be turned an average of six times per year. The manufacturer produces at an average rate of 32 pallets per day. How many pallet positions should you devote to widgets to ensure that widgets turn as required.*

Question 2.9 *A pallet storage facility holds about 10,000 pallets in storage. Arrivals and departures are handled by 7 forklift operators and the average fork-lift travel time from receiving to a storage location and then to shipping is about 6 minutes. Estimate the inventory turns per year. Assume each driver works 8 hours per day for 250 days of the year.*

Chapter 3

Warehouse operations

A warehouse reorganizes and repackages product. Product typically arrives packaged on a larger scale and leaves packaged on a smaller scale. In other words, an important function of this warehouse is to break down large chunks of product and redistribute it in smaller quantities. For example, some skus may arrive from the vendor or manufacturer in pallet quantities but be shipped out to customers in case quantities; other skus may arrive as cases but be shipped out as eaches; and some very fast-moving skus may arrive as pallets and be shipped out as eaches.

In such an environment the downstream warehouse operations are generally more labor-intensive.

This is still more true when product is handled as eaches. In general, *the smaller the handling unit, the greater the handling cost*. It can require much labor to move 10,000 boxes of paper clips if each box must be handled separately, as they may when, for example, stocking retail stores. Much less labor is required to handle those 10,000 boxes if they are packaged into cases of 48 boxes; and still less labor if those cases are stacked 24 to a pallet.

Even though warehouses can serve quite different ends, most share the same general pattern of material flow. Essentially, they receive bulk shipments, stage them for quick retrieval; then, in response to customer requests, retrieve and sort skus, and ship them out to customers.

The reorganization of product takes place through the following physical processes (Figure 3.1).

- Inbound processes
  - Receiving
  - Put-away
- Outbound processes
  - Order-picking
  - Checking, packing, shipping

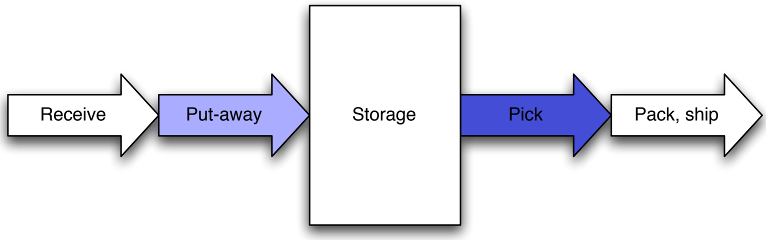

Figure 3.1: Order-picking is the most labor-intensive activity in most warehouses. Travel can be reduced by careful putaway.

A general rule is that product should, as much as possible, flow continuously through this sequence of processes. Each time it is put down means that it must be picked up again sometime later, which is double-handling. When such double-handling is summed over all the tens-of-thousands of skus and hundreds-of-thousands of pieces and/or cases in a warehouse, the cost can be considerable.

Another rule is that product should be scanned at all key decision points to give "total visibility of assets", which enables quick and accurate response to customer demand.

3.1 Receiving

Receiving may begin with advance notification of the arrival of goods. This allows the warehouse to schedule receipt and unloading to coördinate efficiently with other activities within the warehouse. It is not unusual for warehouses to schedule trucks to within 30-minute time windows.

Once the product has arrived, it is unloaded and possibly staged for put away. It is likely to be scanned to register its arrival so that ownership is assumed, payments dispatched, and so that it is known to be available to fulfill customer demand. Product will be inspected and any exceptions noted, such as damage, incorrect counts, wrong descriptions, and so on.

Product typically arrives in larger units, such as pallets, from upstream and so labor requirements are not usually great. (However, mixed pallets may need to be broken out into separate cartons; and loose cartons may need to be palletized for storage.) All-in-all, receiving accounts for only about 10% of operating costs in a typical distribution center [10, 11]—and RFID is expected to further reduce this.

3.2 Put-away

Before product can be put away, an appropriate storage location must be determined. This is very important because where you store the product determines to a large extent how quickly and at what cost you later retrieve it for a customer. This requires managing a second inventory, not of product, but of storage locations. You must know at all times what storage locations are available, how large they are, how much weight they can bear, and so on.

When product is put away, the storage location should also be scanned to record where the product has been placed. This information will subsequently be used to construct efficient pick-lists to guide the order-pickers in retrieving the product for customers.

Put-away can require a fair amount of labor because product may need to be moved considerable distance to its storage location. Put-away typically accounts for about 15% of warehouse operating expenses [10].

3.3 Order-picking

On receipt of a customer order the warehouse must perform checks such as verifying that inventory is available to ship. Then the warehouse must produce pick lists to guide the order-picking. Finally, it must produce any necessary shipping documentation and schedule the order-picking and shipping. These activities are typically accomplished by a *warehouse management system*, a large software system that coördinates the activities of the warehouse. This is all part of the support to expedite the sending of the product to the customer.

Order-picking typically accounts for about 55% of warehouse operating costs; and order-picking itself may be further broken like this [10]:

Activity	% Order-picking time
Traveling	55%
Searching	15%
Extracting	10%
Paperwork and other activities	20%

Notice that traveling comprises the greatest part of the expense of order-picking, which is itself the most expensive part of warehouse operating expenses. Much of the design of the order-picking process is directed to reducing this unproductive time.

In manual order-picking each worker is given a *pick sheet*, which is similar to a shopping list in that it lists the skus to be picked, in what amounts, and where they are to be found. After the customer has submitted his shopping list, the warehouse management system reörganizes the list so that skus appear in the sequence in which they will normally be encountered as the picker moves through the warehouse.

Each entry on the pick sheet is referred to as a *line* or *pick-line* or *order-line* (because it corresponds to one printed line of the sheet). Alternative terminology includes *pick* or *visit* or *request*. Note that a pick (line) may require more than one *grab* if, for example, several items of a sku are to be retrieved for an order.

The most labor-intensive order-picking is the picking of less-than-carton quantities, referred to typically as *broken-case* or *split-case* picking. Broken-case picking is labor-intensive because it requires handling the smallest units of measure in the warehouse and this is generally resistant to automation because of the size and variety of skus to be handled. In contrast, *carton-picking* (picking full cartons) can sometimes be automated because of the relative uniformity of cartons, which are almost always rectangular and packed to resist damage.

The *pick face* is that 2-dimensional surface, the front of storage, from which skus are extracted. This is how the skus are presented to the order picker. In general, the more different skus presented per area of the pick face, the less travel required per pick. An informal measure of this is *sku density*, which counts the number of skus available per unit of area on the pick-face. If a warehouse has a suitably high sku density then it will likely achieve a high value of *pick density*, or number of picks achieved per unit of area on the pick face, and so require less travel per pick.

Sometimes it is useful to interpret the informal measures sku density and pick density as measuring skus or picks per unit of distance along the aisle traveled by an order-picker. One can then talk about, for example, the pick density of an order. An order that is of high pick density does not require much travel per pick and so is expected to be relatively economical to retrieve: we are paying only for the actual cost of retrieval and not for travel. On the other hand, small orders that require visits to widely dispersed locations may be expensive to retrieve because there is more travel per pick.

Pick density depends on the orders and so we cannot know it precisely in advance of order receipt. However, it is generally true that pick density can be improved by ensuring high *sku density*, which is number of skus per foot of travel.

Pick density can be increased, at least locally, by storing the most popular skus together. Then order-pickers can make more picks in a small area, which means less walking.

Another way to increase the pick density is to *batch* orders; that is, have each worker retrieve many orders in one trip. However, this requires that the items be sorted into orders either while picking or else downstream. In the first case, the pickers are slowed down because they must carry a container for each order and they must sort the items as they pick, which is time-consuming and can lead to errors. If the items are sorted downstream, space and labor must be devoted to this additional process. In both cases even more work and space may be required if, in addition, the orders themselves must be sorted to arrive at the trailer in reverse sequence of delivery.

It is generally economic to batch single-line orders. These orders are easy to manage since there is no need to sort while picking and they can frequently be picked directly into a shipping container.

Very large orders can offer similar economies, at least if the skus are small enough so that a single picker can accumulate everything requested. A single worker can pick that order with little walking per pick and with no sortation.

The challenge is to economically pick the orders of intermediate size; that is, more than two pick-lines but too few to sufficiently amortize the cost of walking. Roughly speaking, it is better to batch orders when the costs of work to separate the orders and the costs of additional space are less than the extra walking incurred if orders are not batched. It is almost always better to batch single-line orders because no sortation is required. Very large orders do not need to be batched because they will have sufficient pick density on their own. The problem then is with orders of medium-size.

To sustain order-picking product must also be replenished. Restockers move skus in larger units of measure (cartons, pallets) and so a few restockers can keep many pickers supplied. A rule of thumb is one restocker to every five pickers; but this will depend on the particular patterns of flow.

A restock is more expensive than a pick because the restocker must generally retrieve product from bulk storage and then prepare each pallet or case for picking. For example, he may remove shrink-wrap from a pallet so individual cases can be retrieved; or he may cut individual cases open so individual pieces can be retrieved.

3.3.1 Sharing the work of order-picking

A customer order may be picked entirely by one worker; or by many workers but only one at a time; or by many at once. The appropriate strategy depends on many things, but one of the most important is how quickly must orders flow through the process. For example, if all the orders are known before beginning to pick, then we can plan efficient picking strategies in advance. If, on the other hand, orders arrive in real time and must be picked in time to meet shipping schedules then we have little or no time in which to seek efficiencies.

A general decision to be made is whether a typical order should be picked in serial (by a single worker at a time) or in parallel (by multiple workers at a time). The general trade-off is that picking serially can take longer to complete an order but avoids the complications of coördinating multiple pickers and consolidating their work.

A key statistic is *flow time*: how much time elapses from the arrival of an order into our system until it is loaded onto a truck for shipping? In general, it is good to reduce flow time because that means that orders move quickly through our hands to the customer, which means increased service and responsiveness.

A rough estimate of the total work in an order is the following. Most warehouses track picker productivity and so can report the average picks per person-hour. The inverse of this is the average person-hours per pick and the average work per order is then the average number of pick lines per order times the average person-hours per pick. A rough estimate of the total work to pick the skus for a truck is the sum of the work-contents of all the orders to go on the truck. This estimate now helps determine our design: How should this work be shared?

- If the total work to pick and load a truck is small enough, then one picker may be devoted to an entire truck. This would be a rather low level of activity for a commercial warehouse.
- If the total work to pick and load an order is small enough, then we might repeatedly assign the next available picker to the next waiting order.
- If the orders are large or span distant regions of the warehouse or must flow through the system very quickly we may have to share the work of each order with several, perhaps many, pickers. This ensures that each order is picked quickly; but there is a cost to this: Customers typically insist on *shipment integrity*, which means that they want everything they ordered in as few packages as possible, to reduce their shipping costs and the handling costs they incur on receipt of the product. Consequently, we have to assemble the various pieces of the order that have been picked by different people in different areas of the warehouse; and this additional process is labor-intensive and slow or else automated.
- For warehouses that move a lot of small product for each of many customers, such as those supporting retail stores, order-picking may be organized as an assembly-line: The warehouse is partitioned into zones corresponding to work-stations, pickers are assigned to zones, and workers progressively assemble each order, passing it along from zone to zone.

  Advantages include that the orders emerge in the same sequence they were released, which means you make truck-loading easier by releasing orders in reverse order of delivery. Also, order-pickers tend to concentrate in one part of the warehouse and so are able to take advantage of the learning curve.

  The problem with zone-picking is that it requires all the work of balancing an assembly line: A work-content model and a partition of that work. Typically this is done by an industrial engineer.

Real warehouses tend to use combinations of several of these approaches.

3.4 Checking and packing

Packing can be labor-intensive because each piece of a customer order must be handled; but there is little walking. And because each piece will be handled, this is a convenient time to check that the customer order is complete and accurate. Order accuracy is a key measure of service to the customer, which is, in turn, that on which most businesses compete.

Inaccurate orders not only annoy customers by disrupting their operations, they also generate returns; and returns are expensive to handle (up to ten times the cost of shipping the product out).

One complication of packing is that customers generally prefer to receive all the parts of their order in as few containers as possible because this reduces shipping and handling charges. This means that care must be taken to try to get all the parts of an order to arrive at packing together. Otherwise partial shipments must be staged, waiting completion before packing, or else partial orders must be packaged and sent.

Amazon, the web-based merchant, will likely ship separate packages if you order two books fifteen minutes apart. For them rapid response is essential and so product is never staged. They can ship separate packages because their customers do not mind and Amazon is willing to pay the additional shipping as part of customer service.

Packed product may be scanned to register the availability of a customer order for shipping. This also begins the tracking of the individual containers that are about to leave the warehouse and enter the system of a shipper.

3.5 Shipping

Shipping generally handles larger units than picking, because packing has consolidated the items into fewer containers (cases, pallets). Consequently, there is still less labor here. There may be some walking if product is staged before being loaded into freight carriers.

Product is likely to be staged if it must be loaded in reverse order of delivery or if shipping long distances, when one must work hard to completely fill each trailer. Staging freight creates more work because staged freight must be double-handled.

The trailer is likely to be scanned here to register its departure from the warehouse. In addition, an inventory update may be sent to the customer.

3.6 Summary

Most of the expense in a typical warehouse is in labor; most of that is in order-picking; and most of that is in travel.

3.7 More

Many warehouses also must handle returns, which run about 5% in retail. This will become a major function within any warehouse supporting e-commerce, where returns run 25–30%, comparable to those supporting catalog sales.

Another trend is for warehouses to assume more value-added processing (VAP), which is additional work beyond that of building and shipping customer orders. Typical value-added processing includes the following:

- Ticketing or labeling (For example, New York state requires all items in a pharmacy to be price-labeled and many distributors do this while picking the items in the warehouse.)

- Monogramming or alterations (For example, these services are offered by Lands End, a catalog and e-mail merchant of clothing)
- Repackaging
- Kitting (repackaging items to form a new item)
- Postponement of final assembly, OEM labeling (For example, many manufacturers of computer equipment complete assembly and packaging in the warehouse, as the product is being packaged and shipped.)
- Invoicing

Such work may be pushed on warehouses by manufacturers upstream who want to postpone product differentiation. By postponing product differentiation, upstream distributors, in effect, see more aggregate demand for their (undifferentiated) product. For example, a manufacturer can concentrate on laptop computers rather than on multiple smaller markets, such as laptop computers configured for an English-speaking market and running Windows 2000, those for a German-speaking market and running Linux, and so on. This aggregate demand is easier to forecast because it has less variance (recall the Law of Large Numbers!), which means that less safety stock is required to guarantee service levels.

At the same time value-added processing is pushed back onto the warehouse from retail stores, where it is just too expensive to do. Both land and labor are typically more expensive at the retail outlet and it is preferable to have staff there concentrate on dealing with the customer.