In *Proceedings of the 2000 Industrial Engineering Solutions Conference*, Cleveland, OH, 2000

YOUR DATA AND HOW TO ANALYZE IT

Leon F. McGinnis
Georgia Institute of Technology
School of Industrial and Systems Engineering
Atlanta, GA 3033200205

Steve Mulaik
The Progress Group
100 Roswell Summit, Suite 300
1080 Holcomb Bridge Road
Roswell, GA 30076

Abstract

Data analysis is critical in warehouse design, but often the process of deriving useful information to support good design decision-making is not well understood. This paper will address data and analysis tools in the context of a specific problem—lower-to-mid volume warehouses that are not automated. Computational analysis techniques will be presented, including: 1) how products are partitioned by picking and storage characteristics and then assigned to zones; 2) how storage modes are evaluated and assigned; and 3) how wave/activity planning techniques combined with picking methods are evaluated and selected.

Introduction

Any warehouse design is driven by the requirements to be met by the warehouse. In particular, the warehouse designer must specify the types and amounts of storage capacity, and the methods for assembling customer orders. These design decisions are driven by what the designer understands about the storage requirements, the nature of customer orders, and the rates at which orders must be assembled.

The designer's understanding of these requirements comes from information provided by the client, i.e., the owner or representative of the owner of the warehouse being designed. In this paper, we describe the processing of client information for the case in which the client has a very extensive historical record of customer orders (the order master), and a complete description of the products to be handled in the warehouse (the SKU master). Clearly, not every client has such information; for example, a dot-com startup company typically has no historical data of any type, and only a sketchy idea of the specific nature of future sales. Nevertheless, a clear understanding of how the designer would use a complete historical database provides insight into requirements analysis in the less desirable situation of no data or incomplete data.

For the purposes of this paper, we will make the following assumptions:

1. a *SKU master* file is available, listing each item to be handled in the warehouse, with at least the *SKU ID, SKU description, SKU unit of measure* (pallet, case, each), *weight*, and *dimensions*.
2. an *order master* file is available, listing each *line* of each order for a one-year period; a *line* contains the *order number, ship date, SKU ID, quantity*, and *customer*.
3. data analysis will be performed using Access™ (although any standard database program could be used).

The focus of this paper is on the use of the item master and the order master to support warehouse design, rather than on the design decision making process itself. Also, we make some recommendations regarding preparation for design in the absence of a complete item master and order master.

Preliminaries

A first step in analysis is to *rectify* the data, i.e., to insure that each record in both the SKU master and order master files contains all the required information, and that the information is correct. This can be a challenging task, especially for large databases. While essential and potentially interesting, it will not be addressed here.

The next step in the data analysis will be to identify seasonal items, and group them by season. Seasonality analysis requires an appropriate *statistic* that can be calculated for a SKU and used to assess whether or not it exhibits seasonal demand. One attribute that may be used for seasonality analysis is the *time between orders*, which may be computed as follows. Assume that the order master is sorted by *ship date*. The *time between orders* for a given line (and corresponding SKU) can be calculated by subtracting the *ship date* for the previous line containing the same SKU from the *ship date* for the line being considered. Modern database software allows dates to be converted to numbers and used in arithmetic expressions.

Suppose the *order master* is augmented by adding a *time between orders* field to each line. Then it is a simple matter to compute, for a given SKU, the average *time between orders* and its standard deviation. The *coefficient of variation*--i.e., the standard deviation divided by the average--becomes the statistic to be used to assess seasonality. A value greater than 1.0 is a flag that the corresponding SKU may be a seasonal item. The coefficient of variation (cv) may be added to the records in the SKU master.

Items identified as potentially seasonal may be grouped by computing the average ship date for the corresponding lines in the order master. Note that it may be necessary to exclude "outliers" in computing the average. An outlier would be a line that has a large value for *time between orders,* indicating that it was ordered "out of season."

To simplify the presentation, we will assume for the remainder of the paper that there are no seasonal items. The methods described can be applied easily in the case of seasonal items by grouping them and analyzing them by groups.

Pick-Line Design Analysis

A fundamental issue in warehouse design is whether or not the warehouse will employ a *pick-line*, a special storage area designed to enable more labor-efficient retrieval of the items required for an order. In general, items are good candidates for a pick-line if they are ordered often, and in less than pallet quantities (i.e., cases or eaches). However, since a pick-line consumes floor space, and requires internal replenishment, it involves additional investment and operating expense, so it is not appropriate for all items. Data analysis can help in identifying SKUs that are good candidates for a pick-line.

An *order quantity distribution* can be computed from the order master. For a given order quantity, the distribution function value is the frequency of lines containing that order quantity (which may be a range, e.g., 1 to 4 cases, 5 to 8 cases, etc.). The order quantity distribution provides insight into the potential for a pick-line and the types of pick-line technology that might be appropriate. Based on the insights gained from the order quantity distribution, the designer may choose to partition the SKUs into two or more groups for further analysis.

The order master also can be used to tally the order frequency for each SKU, and the result can be appended to the records in the SKU master. Displaying the *order frequency distribution* will allow the designer to assess whether or not there appears to be a subset of items that have a large order frequency, making them candidates for a pick-line. The SKUs that are candidates for the pick-line, based on order frequency and order size can be selected from the SKU master for further analysis.

The items to be assigned to a pick-line may be determined by sorting the SKU master in order of decreasing order frequency. A *pareto chart* may be helpful in choosing the items to assign to the pick-line.

The storage capacity required in the pick-line is determined by the pick rate of the SKUs assigned to the pick-line and the desired pick-line replenishment frequency. Note that pick rate is different from pick frequency--the former refers to the number of units picked while the latter is the number of times the SKU is ordered. The pick rate (or demand rate) for a SKU may be estimated from the order master simply by adding the *quantity* for all lines containing the SKU.

Suppose the demand rate for a SKU is 300 cases per week, and the desired pick-line replenishment rate is 10 times per week. Then the storage capacity required for that SKU in the pick-line would be 300/10, or 30 cases. Other approaches to setting the capacity for each SKU might be used, for example, assigning equal quantities or equal cubes.

With a set of assigned SKUs and the capacity required for each, the designer may then evaluate alternative pick-line technologies to determine floor space, cost, and throughput. The general approach to data analysis described above permits *parametric design* of the pick-line. The designer can parametrically vary the percentage of SKUs selected using the pareto chart for pick frequency (or pick rate), and also parametrically vary the target pick line replenishment frequency to obtain a range of pick-line capacity requirements.

Wave Planning and Picking Approach

Once the general parameters for the pick-line are set, a designer will then investigate how the orders will be released to the floor for picking. This question will be evaluated usually at the same time that the designer is deciding on what the pick-lists will look like. This first decision is called the Wave Planning and the second is called the Picking Approach decisions.

To make a decision about how the orders will be released to the floor (the wave planning decision) or picked (the picking approach decision), the designer will again explore the order history and SKU information looking for clues. Some of the standard questions asked include:

1. Are there a significant number of single line orders? If so, might we want to pick and process these orders differently from multi-line orders to take advantage of certain productivity opportunities associated with picking single line orders alone?
2. Are most of the orders small in size, i.e. cube? If so, might we want to batch orders together and then multi-order pick those orders into cartons on a cart pushed down the pick-line. This eliminates the packing step and cuts down on travel distance.
3. Are there order cutoff times for different regions of the country that force us to release certain orders together at certain times of the day based on how far away that order must be shipped? If so, do the order lead times force us to consider single order or multi-order picking versus zone or batch picking?
4. Are there a lot of "hot" orders that we will want to release separately and single order pick just to get them out the door quickly?
5. Are there large orders (order cube) that we would want multiple people to be working on at the same time and thus we need to use zone picking to pick these orders?
6. Are there just too many pickers in the pick-line and consequently zone picking is needed to minimize congestion and to preserve picking productivity.

These are typical of the questions investigated at this stage of design, and illustrate why it is so important to have good data from which to work.

Picking Medium Decision

Once a rough listing of the SKUs included in a pick line is available and the decision regarding how they will be picked has been made, a designer will usually then look at the issue of the "picking medium" used within that pick line. The productivity of the pick-line and the accuracy of the picking is heavily influenced by the "picking medium" employed. Examples of different forms of "picking mediums" include paper pick lists, labels, RF terminals, Pick To Light Systems, and Voice Activated picking terminals. Succinctly stated, a picking medium tells a picker what to pick and how much to pick.

The picking medium decision is driven by of the targeted throughput/manhour associated with the pick-line, the desire for picking accuracy, the picking approach used in the pick-line, the experience of the workforce, the density of the picking (picks/foot), the number of picking locations, and the overall throughput of the pick-line in picks/hour. This is an area requiring more research, as generally designers have a bias for certain media in certain situations, and explanations for why one medium is picked over another is often times not based on quantitative analysis but on the designer's prior (successful) experience.

Storage Technology Selection

There are three basic decisions regarding storage systems: (1) what storage technologies to use; (2) how much capacity to provide in each storage technology; and (3) how to zone the storage systems and assign SKUs to zones.

For the scenario addressed here, the basic storage technologies include block stacking, rack systems (single deep, double deep, drive in/drive through), and shelving. The assignment of a SKU to a storage technology will depend in part on the SKU attributes, such as size, weight, stackability, replenishment quantity, and average order quantity.

Fundamental to designing the storage system is estimating the amount of storage space required for a given SKU. There are two basic strategies for capacity estimation. One strategy is to examine the typical inventory in the existing warehouse from which to SKU master and order master were obtained. The typical inventory is then inflated to account for space utilization. For example, in a pallet storage system, if the typical number of pallets in the warehouse is 1500, and the space utilization is 75%, then the capacity of the warehouse is 1500/0.75 = 2000 pallet positions. Clearly, in this approach, the storage system capacity will be directly impacted by the assumption regarding space efficiency, which itself is significantly impacted by the choice of storage technology.

A second strategy for estimating storage capacity requirements is to start from the average demand rate, and assume an average inventory turn-over rate and desired "days of demand" for safety stock.

The capacity required for a given SKU can be estimated by computing the annual demand (total shipped) from the order master, and assuming an *inventory turnover* rate for the SKU. For example, if annual sales are 1800 cases, and we expect inventory to turn over 12 times per year, then the average inventory level is 1800/12, or 150 cases. For dedicated storage, we want the maximum inventory level, which would be estimated as twice the average, or 300 cases. To this we should add an allowance for safety stock, which is a function of replenishment lead time. We might estimate the safety stock in terms of *days of sales* for the SKU, and add 5 days, or 1800*5/360, or 25 cases.

The capacity requirement may be increased or decreased if there is a trend in the shipment of the SKU.

A key decision is whether to use dedicated or shared storage, i.e., whether a SKU has dedicated storage locations, or is assigned to a storage zone which is shared with other SKUs. The advantage of shared storage is space efficiency; the disadvantage is the need to more actively manage the storage space, including the possibility that there is insufficient space available to store an inbound replenishment in its assigned storage zone.

The idea behind shared storage is that not all the items assigned to the zone will be at their maximum inventory levels at the same time. Generally speaking, the more items in the zone, the greater the space efficiency gain from sharing. In the best possible scenario, with all SKUs sharing a single zone, the space required for shared storage would be 50% of that required for dedicated storage. Of course, the best possible scenario is never seen in practice. For planning purposes, a space savings of 5% to 15% might be expected, depending upon the storage technology and the patterns of replenishment and shipping.

Some Conclusions

The availability of a complete and representative item master and order master will reduce the risk and uncertainty in specifying the warehouse design. What should a client do if these databases are not available?

In the discussion above, we've identified a number of ways of analyzing the item master and order master. The analysis provides the designer the following:

- Demand rate
- Turnover
- Days of safety stock
- Quantity per line histogram
- Lines/order histogram
- Lines/week histogram
- Pareto chart for total quantity/week by SKU
- Pareto chart for total lines/week by SKU

If there is no item master or order master, the next best alternative is for the designer and the client to agree on a set of SKU families (perhaps representing seasonality, or product attributes) and then for each SKU family, agree on "typical" values for the items identified in the list above. This structure for the design data allows for examining "what if" questions regarding the future loads on the warehouse.

All of the data analyses described in this paper can be implemented as straightforward applications in Access™. For more information, readers should contact the authors.

Acknowledgments

Professor McGinnis would like to acknowledge funding from the W. M. Keck Foundation, which has made possible the activities and facilities of the Keck Virtual Factory Lab.

Biographical Sketches

Leon F. McGinnis is a professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology, where he leads research addressing industrial logistics. He is Associate Director of the Manufacturing Research Center, founding Director of the Keck Virtual Factory Laboratory, and a Fellow of IIE. Leon may be reached at 404-894-5562 or leon.mcginnis@isye.gatech.edu.

Steve Mulaik is a Director for The Progress Group, a logistics consulting firm in Atlanta, Georgia. Steve is a member of TPG's Management System practice. He may be reached at 770 438 8177 or smulaik@theprogressgroup.com.