

BWH News Letter

Volume 4 Number 4

Why Doesn't My Data Warehouse Work?

Data Warehouse Benefits

Retailers have made considerable investment in their data warehouses. The tantalizing thought of knowing each customer's buying habits seemed to be enough to justify this effort. Who could argue against being better aware of what is going on at store level?

Now that the detail is available, retailers are trying to figure out how to maximize its utility. The real challenge seems to be extracting actionable information from all the details.

Four things are making this effort a challenge; Data Normalization, Causal Data, Knowing What You Want, and having a Basis for Comparison.

Data Normalization

Data normalization is a real challenge in all data warehouses, but particularly for retail organizations because they are not in control of the transaction identifiers they capture.

The Item Identifiers captured by POS scanning systems consist of a combination of UPC's (GTIN's), Industry Standards (e.g. meat, produce, etc.) and locally assigned Price Look Up numbers or PLU's. None of these identifiers provide what is needed to drive assortment or replenishment applications. UPC's are assigned by manufacturers and change anytime there is a physical change in the product. Net weight, packaging, labeling (New and Improved), and formula changes can all result in UPC changes. UPC changes caused by pre-priced or cents-off labeling do affect retailer pricing applications, which must also understand that different labels require a different final price.

This whole concept of data normalization is a variation on the old adage "beauty is in the eye of the beholder". When analyzing data, assortment and replenishment applications are interested in the combined activity surrounding a particular consumer item. All the various net content and labeling variations with different UPC's must have their data summarized for assortment and replenishment decisions.

Causal Data

The classic description of correlation vs. causation is that stock prices and the length of women's skirts usually rise in unison. There is no plausible reason for one to affect the other. The problem with transaction data is that it never tells us why something occurs.

All of a sudden the sales of Brand A sky rocket and the scanning data starts the forecasting systems trying to catch up. It turns out the only other brand is out of stock because they have a DSD driver strike occurring and there is no inventory on the shelf. Understanding why the Brand A sales increased is critical to understanding how to react. Not only should we not expect the sales increase to continue, we need to filter its effect out of next season's projections for both brands.

While working with warehouse out of stock data we identified almost 30 different causes ranging from manufacturers putting an item on allocation because they can't produce it fast enough to poor shelf discipline causing stores to request items which should not be ordered. The only hope to correct a problem is having some understanding of why it is happening.

Transaction systems must allow capture of causal data whenever possible. Feedback mechanisms can allow the human participants to enter ancillary data that helps understanding why things happen. For instance, a receiving application should not only capture the quantity of rejections, but also why and what was done with the rejected product. Open Order files should identify why something was ordered so that receiving systems can route cases to the proper work station for initial processing. The cause for sales transaction discounts should be fully documented by capturing manufacturer coupon and other discounts by line item in the POS transaction logs.

BWH News Letter

Volume 4 Number 4

Knowing What You Want

Sometimes the reason we don't get the benefits from a data warehouse is because we don't know what to ask. Data warehouses capture transaction data but it is the operating applications that need the information that can be interpolated from the data.

An assortment planning application needs to understand things like "how brand conscious are consumers for this category"? What types of consumers shop this category? How does this category respond to price discounts, in store displays, promotional ads, etc? What is the role of the category (Destination, fill-in, base)?

After building a data warehouse to store transactions, businesses need to evaluate their business processes to determine what questions need to be answered. Armed with this information they can establish metrics that allow their operating applications to make intelligent decisions.

Wouldn't it be great if each item had an "importance rating" that was maintained by summarizing the transaction data for each store. Maybe it would be the combination of dollar contribution for the item within the category and the category within the store. It might also take into consideration assortment, so an item becomes more important if there are no other sizes or brands available. It might take into consideration how willing customers are to substitute based on historical sales data showing how much of a price reduction is needed to meaningfully change sales. The type of shoppers that buy an item can be deduced from basket size. The "importance rating" can be used by inventory allocation routines to meter out short supply items or backroom management applications to prioritize pick lists.

By sitting down with the operating managers who are responsible for performance in various areas, a list of critical questions can be developed, then parameterized, then answered from the data warehouse and used by the operating applications to make decisions.

Basis for Comparison

We may be doing better than we realize. Much data can be drawn from within a retailer's organization by comparing the results of various locations. This enables the development of internal expectations so that the managers in various locations can compare their results to their peers. A little competition is good for the soul.

But how does an organization compare itself to the outside world? This is where syndicated data comes into play. Again though, we have the challenge of normalizing the data we get from third parties.

I recall management at one retail organization struggling with warehouse service levels. Their results were constantly worse than numbers they got from an industry survey. It was only when the director of warehousing discovered his peers filtered out the second not shipped that he was able to explain the difference. Rightly or wrongly, his peers had decided that counting the subsequent requests for an out of stock item from the same store before item was shipped again was double counting. I don't know if they ever asked store operations about this calculation.

Feeds of syndicated data need to be evaluated for the calculation of their values and if necessary, adjustments made before they are compared to internal systems.

The Lesson

Direct inquiry into a data warehouse is insufficient for getting maximum benefit. To get a really great return, you need to have a clear understanding of what you want and your operating applications need to have the decision routines that allow them to modify their behavior based on the information you extract from the raw data.