Why Marketing Databases are Not Working... and What to do About It.

John Miglautsch & George Sargent

Database Marketing has been a buzz word for almost a decade.   Some joke that database marketing is selling big computers to companies with lots of money. Companies are spending literally millions of dollars building systems to hold all their data, without addressing the key issues which marketing must face to pull value out of the data. When they are done, they are left with limited capabilities and frustration. Lets take a few minutes to examine the tension between system design and marketing needs. Hopefully, this will help you avoid similar problems when you set off to better understand your customers.

When doing market research or selecting names for mailing, marketing often has difficulty obtaining and analyzing mainframe-based, historical customer and order data. Since the order entry and sales systems that generate the data are crucially important to servicing customers, they generally have been in place for a number of years. The data searching and extraction capabilities of any older systems pale in comparison to state-of-the-art technology. Thus, the aging of the corporate information systems continues to impose frustrating obstacles to effective marketing research. Although that may be the way it is, it need not remain that way.

The Problem: Data Access Issues

In many corporations today (1) much of the data needed by marketing is not available on-line, (2) on-line data is split among multiple computers and therefore difficult to access, and (3) often important data is not even collected.

Limited On-line Data

To determine day-to-day sales, merchandise and marketing can get by with regular printed reports. However, as corporate focus shifts to the customer, it becomes necessary to look at purchase behavior through time. This requires correlating customer orders (the most detailed level) with other customer attributes (such as cluster code or income for consumers and SIC or number of employees for businesses).

To do such trend analysis over time requires vast amounts of historical data. But since operations only needs current data, the historical raw data is usually stored off-line. Some reduced data set is provided in its place which, hopefully, contains the necessary data. Often, this on-line subset is incomplete because marketing's needs change in unpredictable ways.

Unfortunately, the missing data cannot be easily retrieved because it has been archived to the tape library. Given the logistics of working through the IS (Information Systems, formerly data processing) department, the time required, and the cost of the project, often the best alternative is for marketing to make due with the status quo.

Data Split Between Multiple Computers

A second data storage and computer processing problem is that corporations have computer systems which have grown organically over time. Often the needed data is stored on different computers controlled by different departments. This not only complicates data access and retrieval, but the data definitions are often inconsistent making any combined data set less meaningful for decision making.

Data Needed, but not Collected

A final data access problem is that marketing often needs data that the corporation does not collect. In some cases, this is customer or order data that is not needed for traditional operations - but is needed by marketing for their research. Questions like, "How many times would you like us to contact you?" "Do you have other interests or hobbies?" (for consumers) "How many PC's on your network?" "Are there other engineers in your company we should talk to?" (for business-to-business) The data marketing wants changes from month-to-month and year-to-year. IS simply cannot make changes to its stable mainframe order processing systems as fast as marketing's research directions change.

Further, the mainframe computer must be changed by committee. The integrity of the system cannot be violated. Marketing then is in the position of having to prove that these changes are necessary and have economic payback before proceeding. However it is difficult to know the impact of a piece of data you don't have.

In other cases, marketing needs unique environmental data which helps describe customer populations or external market forces. IS generally has no recognized files designed to hold it and cannot easily add it to any existing files. Consequently, marketing must collect it for itself, but then the task of relating it to the other data looms because it is invariably on two different computer systems. Again, the practical alternative is that marketing must make do with a recognizably inadequate data base for its purposes.

THE PROBLEM: USER ACCESS ISSUES

The lack of appropriate data is one kind, but not the only kind of problem marketing faces. Once marketing gets to work with the available data, there are user access issues. (1) marketing personnel have little hands-on access to the data files, (2) the software tools at their disposal are often crude, and (3) the data itself is difficult to understand because it is split among many files that have to be interrelated.

Little Hands-On Access

Most reports produced for marketing are done directly by the IS department rather than by marketing themselves. This is because IS writes the reports to efficiently use expensive and limited computer resources rather than allowing relatively inefficient query tools. Whenever changes are needed in these reports, marketing must have IS do them. This reduces the spontaneity and increases the cost of such changes versus direct "hands-on" action by marketing. Often, by the time the report is "just right" the person requesting it has moved to another position.

Crude Character-Based Query Tools

What little hands-on access marketing personnel gain on the mainframe is generally via monochrome terminals with little or no graphics capabilities. The mainframe rarely presents the user with a "picture" of options to choose from, rather users type a set of commands from memory. Our mainframe-based systems are not "user-friendly."

In the PC world, color is standard and the software supports a mouse to increase user comfort. It has context sensitive help, and best of all, the icons that perform various functions may be grouped in unique ways by the user to customize his or her computer.

When a user moves from their PC to a mainframe to run an application, e.g., the Easy-Trive query tool, they feel as if they have entered an alien world. Users must understand complex data relationships, instruct the computer in how to create joins between files, and constantly check "syntax." Some departments are now asking marketing to learn COBOL. It's not that much worse than the "natural language" mainframe query tool and the programs WILL run fast.

Complex Data File Structures

Another hurdle users face is the inherent complexity of the data tables. If the database is properly structured, IS knows up-front that various problems with data consistency can be headed off before they develop. Thus, it is in their best interest to structure the database properly. Such structuring is a scientific and well-defined procedure. However, in the process, the number of data tables (i.e., files) increase and their inter-relationships become more difficult for the casual user to understand.

The complexity of the files is unavoidable, but the software conceivably could be made to hide some of this complexity from the marketing user, but this is not done with traditional mainframe systems.

THE PROBLEM: PROCESSING AND OUTPUT ISSUES

Available Query Tools are too Slow

Queries need a consistent database from the time they start until they end. If users update the database concurrently with the query, it may obtain inconsistent or erroneous results. Thus, many database systems stop update operations during queries. For example, no order can be entered or changed. While this leads to good technical results, it often raises the ire of the other users and the IS department itself. As a result, marketing is asked to run their queries only on the second Saturday of each month.

Even with more sophisticated relational database systems that allow simultaneous queries and updates, the problem may still exist. Large information systems are expensive and often run near full capacity. They are tuned to handle short requests from many users. When marketing runs lengthy queries they take disproportionate system resources and processing may be agonizingly slow.

The bottom line is that many potentially useful queries are not done and many others must be done after hours. This obviously works against marketing's ability to obtain adequate information.

Iteration

Once marketing finds an interesting result, new questions often arise and they need to experiment further to answer them. For example, they may want to make a slightly different data grouping (e.g., break it down into two year time periods, or get the same data for customers who have ordered in the last six months) and see the chart quickly and automatically reflect the change.

Unfortunately, although this is exactly the kind of capability which would make systems valuable, generally

it is not easily, economically, or quickly done.

Character Based Output

The fundamental issue for marketing is to move from data to strategic decisions. Knowing (for example) how many customers are in a zip code tells almost nothing. That information must be compared to other zips, to the total universe within each zip, (filtered through geo/demographic information), and understood in light of the magnitude of sales campaigns. It is extremely difficult to communicate each step (along with the infinite number of other possibilities).

When the reports do come back from the mainframe, they are character based - containing tables of numbers and statistics. While this is good, it would be better to have such raw data augmented by colored charts graphing relationships between the numbers. This is getting to be essential in communicating complex direct marketing strategies to senior management. PC users may be used to making pretty charts with spreadsheets, but we all know that is not the way it works with the mainframe.

THE SOLUTION

Advances in Computing Hardware

PCs are interactive - that is, when changes are made, the display quickly reflects those changes. Virtually all new PCs are high resolution, color, graphic devices. The power of the PC is increasing. For example, within the last year the fastest PC increased from a 33 MHz Intel 486DX to 50 MHz and 66 MHz 486DX/2 chips. By the end of 1992 the new Intel P5 chip (i.e., the 486 replacement) will run at least two times faster than a 66 MHz 486DX/2.and up to 10 times faster than a 33 MHz 386 ("Late News," PC Week, 8/10/92, p56). The cost of a PC is decreasing - it has fallen 40 percent from June 1991 to June 1992 ("Price Wars Promise No Return to Normalcy," PC Week, 8/10/92, p 56.)

The hot PC software environment is migrating from a character-based DOS environment to a graphics-based Windows' 3.1 environment and in the process is supporting point-and-click operations that generally make it easier to use.

The advanced statistical software which used to require mainframe power has now moved to the PC and added friendly user interface. We all know that PC's are neat (you should see the notebook I'm doing this on) but can they really outperform the mainframe in storing and accessing large volumes of data?

The sizes and fault-tolerance of disk drive systems are rapidly growing. Today multi-Gigabyte, high-performance, fault-tolerant disk systems are available at affordable prices. This allows for the creation of network servers that exceed the performance and storage capacity of the mainframe. (Remember the mainframe must store all corporate data, while marketing's network server need only store marketing relevant data.) Such network servers are available today.

For example, Dell has a 486 based server with a maximum disk capacity of 5 gigabytes, that is capable of supporting over 100 real PCs that are simultaneously accessing a gigabyte database (Catchings, Bill, Van Name, Mark, "What's The Right Size?", Corporate Computing, Vol 1, No. 1, 1992). Other network servers which can be plugged into a PC local area network are available with disk capacities up to 250 gigabytes.

Additionally, database software is available that makes the network system work. The bottom line is that by capitalizing on new desktop technologies, one can have high storage capacity, high performance, fault-tolerant, and affordable systems that outperform the mainframe. They will do so with the user-friendly benefits of PCs because the users' machines are PCs. This is the technology to use when building our new marketing systems.

The Marketing System

The real benefit to the marketing department is not just in switching computers. The benefit is in getting control of the data. Once the data is in a marketing computer, it can be enhanced, modified and manipulated. File layouts can be changed to include outside data, calculated fields and survey information. New technology allows data enhancement on the fly while queries are being processed.

Emerging PC tools like Paradox (from Borland International) can allow you to access many different types of relational databases. Amazing things happen when marketing actually interacts with their own data!

The new marketing system can provide the right information because it has large enough disks to keep more of the raw data. It can wade through massive amounts of information because it is high performance and dedicated to marketing's needs. The power this system delivers to the user can greatly exceed the power per user by the mainframe. When users want to generate a second and third query to explore data relationships, the system should be able to do it in a timely fashion.

The new marketing system can provide hands-on access because it consists of a PC network where users are in control. It can reduce apparent complexities of the database files through innovative graphical displays with point-and-click operations. It can present results graphically because that is the way PCs operate. In fact, several database query programs already do just that.

Finally, it makes the entire process understandable to marketing because it's their hands that run the system and thus they know what they did and will come to understand just exactly what the results represent. This understanding, coupled with better data access and presentation, will provide the ammunition that marketing needs to make convincing recommendations to management about product strategies.

The biggest reason database marketing is floundering is because there is a wide gap between the database and the marketing people. The time has come for marketing to move off the operations system and build a database system of its own. The technology may not be for the faint hearted, but it is well worth the adventure.