Workstation Technology for Marketing Analysis

By Boris Gendelev
Principal, Wheaton Group

Original version of an article that appeared in the January 1990 issue of "Direct Marketing"

[Note: Despite dramatic increases in raw computing power and a proliferation of end-user software tools since the publication of this article, virtually all of the content remains highly relevant. Interestingly, this article recognized the major reasons for, and concepts of, data warehousing before the idea was popularized and gained acceptance. It also anticipated the widespread use of microprocessor workstations for OLAP.]

Analytical workstations: are they for you? If you are having trouble getting a simple analytical report, the answer may be "yes."

Nothing is more valuable to a marketing manager than timely and accurate information. Yet many managers are frustrated, unable to get even a simple analytical report. We believe this is because companies are often slow to accept the solutions new technologies offer.

A spectrum of analytical concepts, such as segmentation, lifetime value, product affinity and channel interaction, have been developed for the purpose of extracting the most relevant information from historical data. Yet, too often, attempts to incorporate these techniques stall, creating friction between marketing and systems professionals. The technological foundation of successful marketing analysis is still a mystery to many managers. Only when its unique aspects are brought to the surface does the need for an entirely new approach become more apparent.

Marketing analysis can be described as a search for actionable information in a sea of historical data, guided by general business concepts and conjectures. It is by nature iterative, driven by the data themselves. Data analysis requires focusing intently; one insight may lead to another, and there is always an urgent need to know if you are on the right track. Yet, as with everything else, there are budgets and deadlines.

To use powerful analytical marketing techniques with consistent success, one must have a foundation of:

- Usable data.
- Flexible data manipulation tools.
- Responsive processing environment.
- Predictable turn around time.
- Low cost computer resources.
- Ability to implement findings.

Contrasted to the traditional approach, this article introduces the evolving technology of microprocessor based analytical workstations. In our view, this technology meets the information needs of marketing management by addressing the fundamentals.

Traditional Approach

Traditional data processing historically evolved around mainframes and, more recently, minicomputers." MIS departments and service bureaus have standardized around a relatively small set of technologies and problem-solving methods. Unfortunately these environments are often unsuitable for marketing analysis tasks:

- The emphasis is on smooth and reliable support of a company's daily operations. Anything else that takes up resources is regarded as a nuisance and given low priority.
- Data files are structured to optimize daily transaction processing. A minimum amount of data is kept on-line.
- Few procedures or incentives exist to ensure correctness of data that are not part of the operational cycle, but which might be vital to marketing (e.g., source and offer codes).
- Developing applications requires mounds of detailed specification and a lengthy formal review process.
- Flexibility, when it is built into the systems, takes the form of extra "buckets" – an approach that only delays obsolescence.
- Personnel, trained to be efficient in a highly structured environment, have difficulty adjusting to midstream changes.
- There is resistance to new, untried technologies.
- Users are charged per unit of processing. This mechanism works well only if processing needs are easy to estimate.
- Software packages, particularly those targeted at high-end marketing users (e.g., MORE/2, and Metaphor) are expensive.

In contrast, the process of marketing data analysis:

- Is impossible to specify completely in advance.
- Often is time critical.
- Demands access to a lot of historical data.
- Requires availability of significant computer resources.

- Can benefit from new technologies such as expert systems and neural networks.

Chart A summarizes the characteristics of traditional data processing and how they impede successful data analysis:

Chart A

	Poor Usability of Data	Lack of Flexible Tools	Low Predic-tability	Poor Respon-siveness	High Cost	Limited Ability to Imp-lement
Competition for resources			X	X		X
Operational file organization	X			X		X
Gaps in data integrity	X					X
Long development cycles		X		X		X
Structured work environment				X		
Resistance to innovation		X				X
High software prices		X			X	
High processing charges					X	X

The conflict between marketing needs and the realities of data processing impacts all stages of data analysis – from data preparation and modeling to implementation.

The need for data preparation itself arises for the most part because:

- The structure of data must be changed to fit marketing perspective. For instance, it is common to find shipping and handling charges, discounts, premiums and taxes as records in a line items file. From an analyst's viewpoint, this information should be on order records.

- Important fields, such as source and offer codes, are missing or incorrect, sometimes in more than 50 percent of cases.

- Duplication is done only on mailing records, not on the main database.

While most of the challenges can be overcome, it often takes a substantial iterative effort, affording little time to develop formal specifications. As early the data preparation stage, the hands-on involvement of a marketing data analyst is needed, but is usually obstructed by the absence of tools that would overcome lack of programming skills.

The statistical modeling itself requires programming, although, with luck, using a high-level language. The desired features of such a language would include:

- Ability to pull together and process as cases all pieces of information pertaining to a customer, such as demographics, promotions, orders, line items and payments.
- Easy ways of creating new customer categories from that rich variety of data.
- Aggregation of statistics across cases by existing and newly created categories.

Unfortunately no popular database query or statistical analysis package (e.g., SQL, FOCUS, SPSS, and SAS) adequately supports all of these functions and most suffer from poor performance. Even if the analyst can program in FORTRAN, COBOL or any other lower-level language, debugging would be very time-consuming, and would have to be done every time something even slightly different is contemplated.

As a result of not having powerful enough data manipulation tools, the data preparation stage is often extended to produce a file easily manipulated by an existing package such as SPSS or SAS. That, however, means committing to premature data aggregation (usually by creating customer summaries and indicators), thereby losing analysis flexibility or having to redo data preparation several times, incurring additional processing charges and making turnaround time unpredictable.

As a rule, the more problems encountered in data preparation, the more difficult it is to implement any kind of segmentation or forecasting in the data processing environment. Tactical analyses, such as R/F/M-based circulation planning or SKU demand forecasting, have a direct impact on a company's data processing because explanatory variables have to be reliably recreated on the main file and tracking mechanisms need to be set in place.

Long-term-oriented analyses such as lifetime value or product affinity usually make their impact on the company more gradually through changes in marketing strategies, making implementation a bit easier.

To summarize, a traditional data processing environment is usually ill suited to meet the needs of marketing data analysis. To expect of it effective and efficient support of marketing management is like expecting a marathon winner to compete and win in sprints.

Microcomputer-Based Analytical Database Approach

Powerful microcomputers are an excellent platform for marketing data analysis, even on a large scale. In fact, a high-end micro has more power potential storage capacity and software than some minicomputers. Moreover, PCs are getting more powerful much faster than mainframes or minis. And all of its resources are usually dedicated to and controlled by a single user. Our experience shows that it is possible to run an R/F/M analysis on a database with half a million customers, one million orders and five million line items in a mere thirty minutes.

The cost per unit of processing on a PC is about one-thirtieth of that on a mainframe and one-tenth of the cost on a typical minicomputer. The cost per unit of disk storage is about half. While hardware costs are falling overall, the relative cost gap is only getting larger.

The notion of performing data analysis in a dedicated computing environment goes hand in hand with the notion of maintaining a comprehensive analytical database with its own set of support programs and manipulation tools. The availability of a wide range of easy-to-use microcomputer software, from template-based generators to expert system shells and object-oriented languages, makes this feasible. The advantages are many:

- The process of its construction will force a systematic review of available data and produce long-term solutions to make it more usable.
- Once constructed it will require only relatively infrequent updates, thus minimizing the dependency on the main system.
- Performance can be optimized to a set of very specific needs.
- Initial investment in equipment is low and incremental processing costs are close to zero.
- With the help of specially tailored tools, new analytical ideas can be tried immediately, without involving data processing.
- Many interrelated analyses can be performed on the same consistent base.
- Customer mailing selection and keycoding can be easily preformed from such databases, integrating these tasks with analysis.

Chart B summarizes benefits of the microcomputer workstation approach in relation to the needs of successful analysis:

Chart B

	Usability of Data	Flexi-bility of Tools	Predic-tability	Respon-siveness	Low Cost	Ability to Imp-lement
Cleaned up data	X					X
Single base for analysis	X		X		X	X
Isolation from main data processing			X	X		
Optimized performance				X		
Low equipment cost				X	X	
Zero processing cost					X	
Specialized software		X		X		
Integration with mailings						X

This new approach is counter to recent trends toward closer systems integration. However, we have to remember that a marketing analysis system, while sharing terminology with a marketing tracking system, is different because it serves the planning rather than control function in a company. It would benefit little from being embedded into one super system along with fulfillment, inventory control and accounting. In most cases, a selection and transfer of customer records would be a satisfactory level of interface to a mailing system. While other systems can be impacted by results of data analysis, periodic feedback is sufficient.

The success of the new approach stems from matching appropriate technologies to the unique nature of marketing analysis. To be manageable, analytical databases must be organized with minimal redundancy and maximum useful transaction detail. Experience shows that the marketing analysis database will be just one-twentieth to one-fifth the size of the company's main database.

A well-conceived database construction sets the stage for fast economical analysis by putting the data into a ready to use form. The task of restructuring, reforming, rematching and recoding the data should now be viewed as building a permanent interface between the data processing function and the analytical database. It is best done by someone outside the data processing department, who appreciates marketing analysis needs and has good computer skills.

Once the analytical database is constructed, flexible manipulation of the data can be accomplished by constructing specialized high-level tools that have the following functions:

- Aggregation of counts and amounts by one or more categories in the form of frequencies, cross-tabs and means. For example, quarterly demand and orders by source and data of first purchase.

- Ability to apply statistical procedures such as regression.

- Flexibility for a general programming language for specifying transformations of variables, such as recoding or categorization.

- Support for multiple record types (e.g., customer, order, item) within each customer case. This means an ability to easily construct new categories by formulating questions such as:
    - "How much did a customer spend last year?"
    - "How many times did a customer buy in given product category?"
    - "What was the main product category of a customer's first order?"
    - "What product families did a customer purchase from last year?"
    - "What combinations of product families does a customer buy from?"
    - "Was a recently bought product purchased before?"

- Access to reference tables. This might be needed for assignment of product families based on SKU, for instance.

- Easy interface to other software such as spreadsheets, report generators and presentation graphics.

- Ability to create mailing selections.

These capabilities can be either developed or acquired at a cost often substantially lower than creating a comparable functionality in a traditional data processing environment.

From our experience in managing analytical databases we can report that the appetite for information grows with the ability to access it easily and inexpensively. One apparel cataloger at first wanted to know lifetime customer activity patterns by source, then also by size and type of the first transaction. Recently, queries evolved to include: "Do customers buy tops and bottoms together and what color combinations do they buy?" "Is there a consistent preference for a color?" "What is the long-term effect of backorders and unfulfilled orders?" What began as a wish list is now an integral part of their marketing decision process.

In planning, strategic or tactical, information is everything. Marketing analysis is a tool of planning. In a climate of increased competition and rising costs, one cannot afford to over1ook new and promising information technology. If your corporate goals include smarter customer acquisition and development, improved efficiency of promotions, better pricing and coordination of marketing channels, do not overlook analytical workstations.

Boris Gendelev is a Principal at Wheaton Group, and can be reached at 847-205-0916 or boris.gendelev@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts.