**Database Design is Difficult**

[*Larry Goldman*](javascript:void(0);). [**DM Review**](http://proquest.umi.com/pqdweb?RQT=318&pmid=56564&TS=1256896408&clientId=2606&VInst=PROD&VName=PQD&VType=PQD). New York: [Jan 2008](http://proquest.umi.com/pqdweb?RQT=572&VType=PQD&VName=PQD&VInst=PROD&pmid=56564&pcid=37987691&SrchMode=3). Vol. 18, Iss. 1; pg. 24

**Abstract (Summary)**

Database designers must predict difficult queries from the beginning. Scenario and query prototyping should be standard operating procedure so you don't find out during testing that your data model can't support the queries. Hardware and software changes and modifications are not the answer. We have seen not-so- sophisticated marketing departments bringing large-scale IBM and Teradata systems to their knees with straightforward data models.

**»**  [Jump to indexing (document details)](http://proquest.umi.com/pqdweb?did=1406318761&Fmt=3&clientId=2606&RQT=309&VName=PQD#indexing)

|  |
| --- |
| **Full Text**   (825  words) |

*(Copyright c 2008 Thomson Media. All Rights Reserved.)*

I regularly hear the complaint, "I can't get basic information" from the business side - particularly from sales and marketing individuals. Often, the phrase is distorted. Sales departments typically have access but haven't spent the time to learn how to download the report or use the cube. Often, they have forgotten their training, don't remember where or how to access the reports and have given up. Marketing departments can usually get the basic information, but "simple" or "basic" is often defined differently between the customer intelligence team and marketing.

Customer analysis is very difficult. Many relational databases and database designs are not able to handle some of the difficult concepts supported by customer analysis. Time series, product correlation and the nonexistence/existence of activities make analysis very difficult. Many times these requirements come across as ad hoc requirements that may be ignored in order to optimize for scheduled management reports.

To avoid being blindsided by these requirements and requests after implementation, those responsible for marketing databases should consider the following types of analysis:

\* Multiproduct or service relationship to the customer.

\* Identifying specific events that have occurred over time.

\* The nonexistence of events, transactions or behavior within the database.

Product Correlations

Though all sales and marketing databases are customer-centric at their core, the first question from a customer analyst is, "What did they buy?" This simple question is usually easy to address, as is the management report that shows sales, quantities and profitability broken down by product category or division. The more difficult questions include:

\* How many product categories do specific customers use or purchase?

\* Which product categories do specific customers use or purchase?

\* What combination of product categories do specific customers use or purchase?

\* How many customers use a specific combination of products?

\* What type of customers use specific combinations of products?

Many organizations focus on the breadth of their product lines. Typically, there is a huge opportunity to drive more usage across product lines rather than push single products to individual customers. This is true for telecommunications companies who want their customers to use landline, wireless and DSL or technology companies who want to sell computers, printers, modems and other accessories.

Time-series events and time series help us understand how fast customers adopt or respond to new marketing pitches. Common questions include:

\* How fast did the customer get to a certain revenue level?

\* How long did it take the customer to buy their second product?

\* How long has it been since the customer's last purchase?

\* What is the average time between purchases?

\* After introducing a marketing program into the field, how long did it take customers to adopt?

Like product relationships, it is difficult for databases and query tools to compare dates in between records in an efficient manner. It can also be difficult to know which record to compare to in a high-transaction environment. Clickstream analysis poses this problem. Correlating browsing to buying is a very disconnected process. It is difficult for query tools to correlate massive amounts of clickstream data together without help from the database.

All dates need to be documented across transaction and customer records. Last purchase date, last login date and last click through date should be precalculated instead of calculated on the fly. You may find yourself attaching dates to a high percentage of fields in your customer record. It is just as important to know when a customer purchased from a specific category as the last time they purchased from that specific product category.

Nonexistence of Activities

As already stated, it is hard enough to identify specific events in the database. It becomes even more difficult to identify what the customer has not done yet. For example:

\* Who has purchased product category A but not category B?

\* Who has not responded to any of the cross-sell campaigns?

\* Who has not logged into the Web site in a while?

These types of queries make it very difficult to discern inaction - which is just as important as actions that have happened.

Database designers must predict difficult queries from the beginning. Scenario and query prototyping should be standard operating procedure so you don't find out during testing that your data model can't support the queries. Hardware and software changes and modifications are not the answer. We have seen not-so- sophisticated marketing departments bringing large-scale IBM and Teradata systems to their knees with straightforward data models. To avoid the "I can't get basic information" complaint, you must simulate these difficult queries and work with the business on how they will approach certain list selections or analysis. You can't defend against creative marketers because they will always find the killer query. Just try to make sure everyone agrees on "killer" versus "basic."

Larry Goldman is president of AmberLeaf. He may be reached at larry@amberleaf.net.

(c) 2008 DM Review and SourceMedia, Inc. All Rights Reserved. http://www.dmreview.com http://www.sourcemedia.com