

Build a High-Performance Data Warehouse to Handle Customer Relationship Management Applications

By Craig Abramson



The old neighborhood retailer knew the secret to successful CRM long before it became one of the hottest concepts for marketing in the late 90s: know as much as possible about your customers and market to their needs in order to develop stronger relationships with them. The data collection and marketing techniques in today's information economy have changed from the old days, but the basics remain very much the same. CRM initiatives are designed to strengthen and expand ties with customers by tapping the full potential of a company's existing customer base. When implemented effectively, they can help a company provide customers with relevant choices, build loyalty, market new services to existing customers, and create customer-centric strategies.

The three areas that a CRM program typically focuses on are sales, customer service and marketing automation. Sales includes everything from field sales and call center telephone sales to retail and e-commerce. Customer service encompasses field service and dispatch technicians, Internet-based service, and call centers. The third area, marketing automation, includes data-cleansing tools, data analysis and business intelligence tools, content-management applications, and a campaign management system. (Robinson, 2000) According to Gartner

Dataquest, revenue in the area of marketing automation in 2002 is expected to reach \$25.3 billion and grow to \$47 billion by 2006. (Rohde, 2002)

This growth in revenue demonstrates the importance that companies are placing on their CRM initiatives. But with all of the hardware and software choices that must be made, it's very easy to get bogged down in the decision-making process. Although the idea of developing a program may seem like a daunting task, there are steps a company can take to ensure a smoother implementation.

The Data Warehouse

The foundation for CRM processing is the data warehouse. It contains the integrated, historical, granular data that can be analyzed to determine where customers live, their age, what they're purchasing and more. (Inmon, 2001) The data warehouse helps speed up querying time so a company can quickly determine the latest trends. It also allows a company to manage information more efficiently and store data for longer periods of time at minimal cost. Data is much easier to access because it's now located in a single database instead of scattered throughout various systems. And most importantly, a data warehouse can provide vast improvements to querying and reporting, helping a company uncover a wealth of information that can be used to improve its CRM.

A Quick Guide to Building the Data Warehouse for CRM

After a company makes the decision to build a data warehouse for CRM, they'll want to do the job as cost effectively and easily as possible. According to Bill

Inmon, considered to be the “father of the data warehousing,” the data warehouse should be built iteratively, in small, fast bursts of development. The data warehouse developer should quickly move through the design and development steps even without having the complete processing and data requirements. These steps begin with the data model, which consists of both a high-level and mid-level model. The high-level data model contains the company’s major subject areas and is completed before the first iteration of data warehouse development. It is not expected to take much time to complete this model since it is considered a routine job. The mid-level data model includes keys and attributes, and their grouping together. Only the parts of this model that coincide with the first iteration of the data warehouse have to be completed.

(Inmon, 2000)

Of course, the first iteration also requires a selection of software and hardware. The hardware is usually mainframe, parallel, or client/server. The software is usually either full function DBMS (database management system) or specialized software designed to improve the performance of the data warehouse. The criteria for selecting this technology should include such factors as the volume of data to be used, the speed at which a company needs the information, what kind of analysis is going to be done, and how much the technology costs. (Inmon, 2000)

Another key criteria for selection is scalability. A company should anticipate that the number of users of the data warehouse will grow and that the amount of data will also increase. This should be considered when making decisions on such things as processor capacity of the database servers, quantity and configuration of the initial memory, and number and type of disk drives. (G.A. Sullivan, 1998)

Once the hardware and software are purchased, the first iteration can begin. The amount of data to be loaded during this process should be large enough that it contains meaningful results but small enough so that it can be done quickly. The next step is to select an entire subject area for implementation and then add the physical attributes and unit of time for the data. Once this is accomplished, the system of record needs to be identified. This is the source data that resides in the operational, legacy systems environment and is the most complete and accurate. The data is transformed and metadata is created. Space is allocated and the data is then loaded into the warehouse where it can be analyzed. (Inmon, 2000)

Although building the data warehouse is the foundation, organizations worldwide are discovering that achieving the goals of CRM also requires a firm base of quality data.

The Dangers of Bad Data

Here are some of the consequences that companies face when their data is not of the highest quality:

- A manufacturer sold off what it thought was excess stock because of invalid data. The company was actually short of stock, leading to thousands of unfilled orders, unhappy customers, and lost revenue.
- Inaccurate data caused an insurance company to raise its risk exposure too high and suffer very expensive losses on many of the policies it wrote.
- Employment data that was not updated on time caused a government agency to request 25% fewer staff members than it needed. The result was a substantial layoff of staffers and excessive overtime and poor morale among the remaining staff members.

All of these problems could have been avoided if more attention had been paid to data quality.

Variant Data Invalidates Intelligence

The quality dilemma is easy to see in a simple example of invalid data. Here are five variations that might appear in the key field for a data warehouse query:

Jon Smith

Jonathan Smith

J. Smith

Jon R. Smith

Jonathan R. Smith

Although it is unlikely that these variants will occur all at once, even two or three can severely skew results. Variants can cause overestimates if duplicate records have been created, or underestimates if a substantial number of records with variants such as misspellings or abbreviations are missed during aggregation.

Three-Step Data Quality Plan

High quality data is defined as data that is complete, valid, consistent, timely, and accurate. Following is a three-step procedure that will help make sure the data is of the highest quality:

1. Research
2. Remediate
3. Enhance

Although the third step is optional, enhancement is important because it can improve query results in a variety of ways, especially for marketing and sales applications.

Research

The first step, research, has two phases:

1. Identify any inconsistent data.
2. Find out how the bad data is entering the system.

A variety of tools can be used to identify bad data. Some of these tools are dedicated and often quite slow while others are fast and flexible. It is also important to know how bad data is, especially if data cleaning procedures were

followed rigorously. Some possible sources of bad data are poor training of data entry personnel and/or inadequate or missing cleaning programs. For example, keys that are sometimes in mixed upper and lower case and at other times all in upper case could be the problem if a cleaning program is not case sensitive but the query program is.

Remediate

Once the variant data has been identified, a procedure to remediate that data and to prevent future corruption must be worked out and put in place. Again a dedicated tool can be used for this procedure.

Enhance

A thorough data quality program includes two phases. First, high quality data is prepared for loading. Data is either captured in a standard, error-proof way or it is cleaned in preparation for loading, as has been discussed above. Next, data is enhanced with demographic and/or lifestyle information before the actual load.

Enhancing customer data usually entails combining multiple sources of data, and this data is often held in multiple databases on disparate platforms. Using a data manipulation tool makes coordinating all these different sources much easier.

Building Relationships with Customers

One company that has implemented a successful CRM strategy is Wachovia Corporation. They realized that every day, Americans opened their mailboxes to find applications for financial products from banks they often knew nothing about. And many promptly threw those applications into the wastebasket. That's what Wachovia wanted to avoid. An interstate bank holding company, Wachovia offers credit and deposit services plus insurance, investment and trust products to consumers in the southeastern U.S. through its network of retail offices and ATMs as well as telephone and Internet banking. Wachovia also serves customers nationwide through its credit card business. Not content just to grow the number of accounts, Wachovia's strategy centers on building relationships with existing customers so they bring repeat business and referrals. The trick is to market the right products to the right customers based on good information and advanced decision support tools.

"Most banks simply buy name-and-address lists and send out millions of applications. They don't know the recipients," said John Collins, Wachovia's manager of data warehouse architecture. "If they get one tenth of one percent return, it's considered a successful mailing. That kind of approach takes a lot of people, processing, postage and paper. We do not want to expend those mass mailing resources if, by doing some homework, we can market to clients who will respond favorably to specific products."

To do this, Wachovia has built an enterprise-wide data warehouse. Running on DB2 in a distributed UNIX (AIX) environment, the warehouse contains a wealth of historic information about clients, their households, their account activity and their relationship with the banks. Using advanced filtering algorithms, Wachovia can segment its market and spot product trends. Results from a preliminary system indicate that Wachovia can get a 30 percent response rate for some new services by focusing its marketing efforts.

Warehouse Processing Far Exceeds Mainframe

To maximize its return on investment in its data warehouse, Wachovia needs to process the greatest amount of data in the least amount of time. "You only have so many hours in a day to process all the data that results from your customer relationships," observed Collins. Processing the bank's millions of daily transactions overnight on its mainframe systems is enough of a challenge, but the processing load for the warehouse is even greater. That is because the warehouse works with vast amounts of historical data.

Warehouse processing, in Collins' experience, "is often tens, hundreds, thousands times larger than in the mainframe environment. On our mainframe we're thinking about the current balance or the last transaction. In our warehouse environment, on the other hand, the biggest dimension is historical." Collins deals with time-related data such as customer weekly bank balances over

a year, household income over 10 years, and changes over long periods in marital status, residence and account status – all multiplied by millions of clients.

Accelerating Database Loading Is Critical

To handle the huge warehouse-processing load, Wachovia exploits RS/6000 SP technology, clustering multiprocessing nodes in a parallel environment. In order to succeed in this warehouse environment, it's critical to reduce the elapsed time of the database loading process. Collins said, "We process data in parallel. Instead of one serial program that processes multiple gigabytes of data, we have 10 or 20 copies that split the data into manageable units and process it simultaneously."

In designing the new platform, Collins knew that in order to accelerate the data warehousing application and database loading, he would have to use a solution that would provide high-performance data preprocessing. The solution chosen was SyncSort UNIX from Syncsort Inc., in Woodcliff Lake, NJ. "When we started building this warehouse," Collins recalled, "the monthly loads of 50-70 gigabytes of historical data were taking us two to three weeks to process. Using SyncSort for preprocessing, we've brought that down to five days, or even four in some situations."

Data Warehouse Provides Competitive Advantage

Wachovia expects the warehouse to become the single source of client decision support information for the bank. The warehouse will have complete information

for every customer, whether that customer has a checking account, a trust account or credit card. Currently, these different data applications maintain their own historical data in their own master files.

Collins believes the warehouse is critical for Wachovia's future success. "I believe that companies who don't have a data warehouse or advanced decision support plan will lose competitive advantage. The financial services industry is moving toward extremely customer-focused marketing. The people who work with customers need to know more about them than just their account numbers. The data warehouse also allows you to spot trends, to see, for example, that one type of product is really taking off or that another product is taking off but only to a certain type of individual."

It's Collins' job to ensure that users of Wachovia's data warehouse have the right technical tools to mine the data effectively. "The UNIX platforms in general are not as mature as the mainframe," he observed, "but if you're running a large portion of your business on a platform like this, you need components that are rock solid."

A Final Piece: A Backup and Recovery Strategy

Wachovia has demonstrated how careful planning and detailed analysis can lead to a very successful CRM program. But the program depends on the availability

of the data so it's also critical to implement a backup and recovery strategy.

When choosing a backup and recovery tool, there are several factors to consider:

- Speed of backups – determine what your backup window is and then test a product to see if it can perform in that timeframe. One way to increase the speed of the backup is to select a tool that performs differential or incremental backups.
- The speed of the restore – getting data back as fast as possible is what really matters. Determine how quickly a backup solution will be able to restore your files.
- How easy it is to restore a file – does the tool use a graphical user interface (GUI) that almost anyone can operate or are complex commands needed?
- Compatibility – a high-performance backup must be able to handle NetWare, UNIX and Windows. This compatibility also allows for scalability.
- Security – a backup solution should have various levels of encryption and device distribution to protect the data as it travels across a network.

Summary

When implemented effectively, a CRM program can live up to its promise of stronger customer relationships and loyalty. Some of the key steps involved in developing the program include choosing the right hardware and software, building the data warehouse as quickly as possible, improving the quality of the data and reducing the elapsed time of the database loading process, and having a backup and restore strategy. Wachovia is just one example of a company that has implemented a successful CRM initiative, helping to strengthen and expand

ties with their customers. They were able to tap the full potential of their existing customer base. But as they demonstrated, it all begins with the data warehouse and data quality.

References

G.A. Sullivan. "Building a Data Warehouse for Improved Decision Making"
<http://www.gasullivan.com>, (June, 1998)

Inmon, William H. "Building the Data Warehouse: Getting Started,"
BILLINMON.COM, (2000), 2-14

Inmon, William H. "CRM and the Corporate Information Factory,"
BILLINMON.COM, (November, 2001), 6

Robinson, Robin A. "Customer Relationship Management"
<http://www.computerworld.com>, (February, 2000)

Rohde, Laura. "Gartner: CRM Revenues to Rise 15 Percent this Year"
<http://www.itworld.com>, (April, 2002)

About the Author

Craig Abramson is a technical analyst at Syncsort Incorporated, focusing on the latest data warehouse performance solutions. He has over seven years experience in the field working on projects dealing with data warehousing, database management and Web log processing.