A BTI Case Study

Smart Recommendations for Small Businesses at Network Solutions®

January 2011

"We saw the growing need to have a dynamic storefront that provides more personalized experience to our website visitors and existing customers. The recommendation engine provides customized offers and educational advertisements on our storefront. It is the key to help our website visitors and existing customers to pick the right product and service for them "

Shiv Verma, Director
Business Intelligence & Reporting

This is the story of the Recommendation Engine project at Network Solutions, a pioneering service provider in the use of the Internet by business. This project developed an application that segments website visitors by their needs and then feeds this information about these segments into a Recommendation Engine. The story focuses upon the use of real-time analysis of integrated data for enabling visitors to learn about relevant Internet services, such as web hosting and search engine optimization.

Company Background

Network Solutions, LLC, is a leading provider of on-line services for business. Founded in 1979, the company is a pioneer in providing Internet services with more than thirty years of experience. Network Solutions manages more than 7 million domain names, over 1.5 million email boxes and over 350,000 websites, with customers in about 200 countries. Network Solutions provides on-line support services including design, security, and search engine optimization. The product portfolio spans domain name registration, web site design and hosting, e-commerce solutions, online security products, and search engine marketing ("SEM") and search engine optimization ("SEO").

Network Solutions is privately held by General Atlantic, a leading global growth equity firm providing capital and strategic support for growth companies with the expertise of a collaborative global approach, sector-specific expertise, long-term investment horizon and a deep understanding of growth drivers.[1]

Corporate Strategy

The focus of Network Solutions is to enable small business to be successful in a global online economy by making the Internet simple and attainable, as illustrated in the following quote:

[Network Solutions] makes it simple and affordable for our customers to build and manage an on-line presence. While we're a technical services company, we specialize in supporting those who aren't technically savvy …so that they can realize their dreams of self-reliance, entrepreneurship, creativity, and financial independence.[2]

Network Solutions works to level the playing field for small businesses, so that they can leverage the Web to gain advantages that previously only larger companies enjoyed.

Business Problem

Network Solutions has to maintain a viable presence in the Internet service market amid rapid technology changes and global competition. While being a thirty-year pioneer in this market, Network Solutions needed to be more analytic to increase the effectiveness with its online marketing across its core web property ("Storefront").

The current home page for the Network Solutions Storefront is shown in Figure 1below. At the top, note the layout of audiences (Business & Professional, Web Professionals) and services (domain names, websites). The bottom left of the home page shows its key service offering of domain name registration.

Figure 1–Current Home Page for Network Solutions

After reviewing the best practices of eBay, Amazon, and other leading websites, Network Solutions realized that it had access to the data for improved marketing but were not fully utilizing it, particularly for analysis like propensity modeling.[3] Although there was a monthly analysis of orders that created forecasts, these reports were not used for aggressive targeted marketing.

Business Solution

As a solution to improve its online marketing, Network Solutions initiated a project to develop a home grown Recommendation Engine to support dynamic merchandizing and, thereby, to enhance the user experience when visiting the Storefront. Network Solutions needed to classify its customers into segments possessing similar needs and then to generate recommendations for each segment as customers in that segment visited pages of its Storefront. If Network Solutions made highly relevant recommendations, Storefront visitors would have a personalized experience, seeing relevant offers and not bothered by other offers, with the goal that customers would be more likely to purchase services and more likely revisit the Network Solutions website.

At that time, Network Solutions broaden its focus for Teradata into operational intelligence for dynamic merchandizing and for customer service. Navin Ganeshan, Director Product Strategy, described the issue as:

> *How do we use all this intelligence operationally to directly drive action by our staff or to influence actions by the customer? Supplying intelligence to customer marketing has become very complex given all the alternatives for up-selling and cross-selling opportunities versus the value to the company. Network Solutions needed to digest all that information so that the sales rep could focus on the most relevant aspects to the customer.*

The goals were to increase the propensity for a customer to purchase additional services and to increase the total revenue from those purchases. A key factor was propensity, which is the probability of accept an offer given a set of prior conditions.

As the team lead on the Recommendation Engine project, Sathish Mallamula, Senior Manager of Business Intelligence gave an overview of their solution.

> *We examine various website visitor data points to determine what they have been looking for previously. We collect web server logs and operational systems data. Through nightly feeds we put all the data into the data warehouse along with a small number of real-time feeds. The results are used to drive the dynamics of our portal so we enhance the customer experience. For instance, first-time visitors get special discounts to get them engaged with Network Solutions.*

Network Solutions has a rich product set requiring its Storefront to have carefully designed information architecture. It has simplified several different paths through the web site into few principal channels, which has increased the effectiveness of the Recommendation Engine. Network Solutions upgraded its simple ad engine to incorporate the Recommendation Engine's output, which greatly improved the ad serving capabilities.

Network Solutions approached dynamic merchandizing incrementally. The bottom right area of the website held a rotating placement five selected ads out of a set of ten. The selected ads were based on all the warehouse data, such as online behaviors, product owned, interest in other products, and so on. Based on the positive results, Network Solutions expanded the use of this capability from this one area to all areas of the Storefront, using the same mechanism.

The targeted, rotating placement of ads on existing web pages enhanced click-through rates and revenue generation. An example is shown in Figure 2 with the dynamic ads circled.
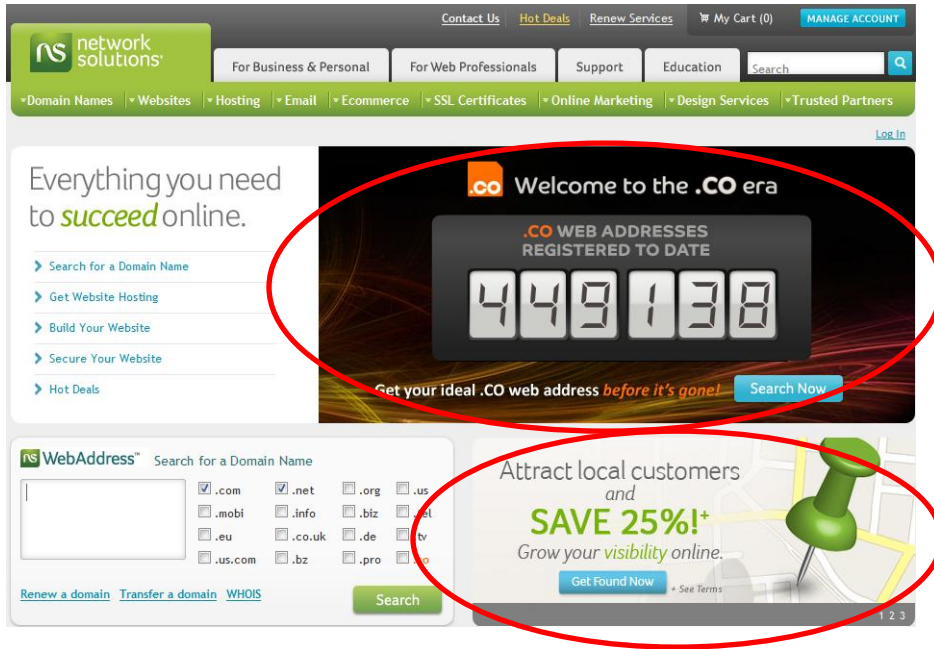
Figure 2 - Dynamic Ads Internal to Web Pages

While simple in its end-user manifestation, the delivery of these targeted ads incorporated a wide variety of data sources and software platforms. The technical architecture for the solution is shown in Figure 3.
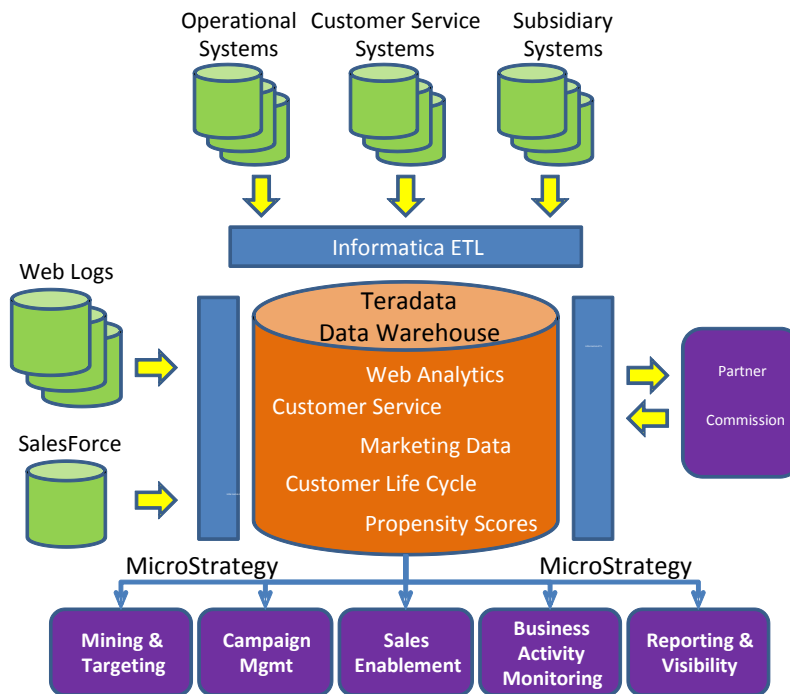
Figure 3 - Architecture Flows and Tools

Network Solutions uses Web Log DB[4] to collect and clean the raw web logs from the Storefront websites. An Extract-Transform-Load (ETL) process incorporates the web logs into the Teradata Data Warehouse. This ETL process is performed nightly starting just after midnight (US Eastern). The numerous web servers feed the Informatica ETL servers that load data into the Teradata System. The ETL processes utilize many source-to-

destination mappings, all managed by Informatica. These web logs provide all of the thousands of clicks made by visitors to the web site.

For instance, one of several product lines is the registration of internet domain names (such as "mybusiness.com" or "myfamily.org"). To understand what visitors seek, the web server log data is captured by the ETL processes and loaded into the Teradata Data Warehouse. For example, if a customer submits a domain name search and the domain name is not purchased, Network Solutions will use this domain name in a future campaign to make an offer when the visitor returns.

BI Storefront Recommendation Model

Once all the web and application data is loaded into the data warehouse, various dependent data marts are built in batch mode. In particular, the BI Storefront Recommendation Model (SRM) is a data mart containing many columns. It contains various keys for customers, along with their summarized customer profile data. For instance, anything in which a customer expresses interest but was not purchased is captured in a "wish list" table.

The major part of the Recommendation Engine is a set of database tables that store the recommendations and propensity scores by Customer Unique Identifier. The Recommendation Engine uses various rules-based processes that operate on these tables and also performs considerable segmentation analysis as the recommendations are built.

The SRM updates only at night. Hence, recommendations are based on customer interactions from the previous day and prior, although there is in-session data used for placing ads. In-session is the real-time aspect of the recommendation that will mix-n-match the recommendations from SRM before rendering the most relevant content.

The visitor could be a customer (who previously registered with Network Solutions) or a visitor (who has not registered). When a user visits a Network Solutions site, a persistent cookie[5] is sent to the client browser. This cookie is used initially to identify an anonymous visitor, and to tie together previous interactions of visitors to dynamically build the web pages, even if anonymous. Once a person signs in with Network Solutions, the Recommendation Engine uses the Customer Unique Identifier to access preferences.

Content Management System

The SRM is used to refresh an operational data store (ODS) for the Storefront. The content within the ODS and SRM is almost identical. (The primary reason for the ODS is to handle the sub-second clicks on the websites, which amounts to more than a million clicks per day.) In addition, a separate Content Management System (CMS) contains the ads and ad copy for dynamic display on the webpage. Using web services, CMS merges its rules with the rules in the ODS to generate the dynamic HTML pages to display to visitors and customers. The Storefront will read the cookie and display the appropriate ads/banners.

One unique aspect of the Recommendations Engine is the rapid changes that can be made. New data elements, rules, and processes can be added or changed quickly. For a major new campaign with its collection of rules, the required changes can be made to the data warehouse in a week. Campaigns always use a unique code for a landing webpage. The applications can then track how many visitors are clicking that landing code and which of those convert to a sale. There are always new landing codes being added to uniquely identify a campaign, promotion, or activity. A new flag – landing code – in the SRM tables can be implemented within a few days and linked to the CMS rules.

Customer segmentation also enables different Storefront treatments for customers in different geographies, purchasing behaviors, or product genres. For example, based on the IP-address such as Atlanta, the CMS rules might suggest, "Here is the campaign for that region".

Storefront Ad Performance Dashboard

The Storefront Ad Performance Dashboard (SAPD) was implemented using MicroStrategy products to help track how the recommendations-driven ads were performing. Figure 4 shows an overview report on ad performance by date, while Figure 5 shows a more detail view of ad performance over the past four weeks.
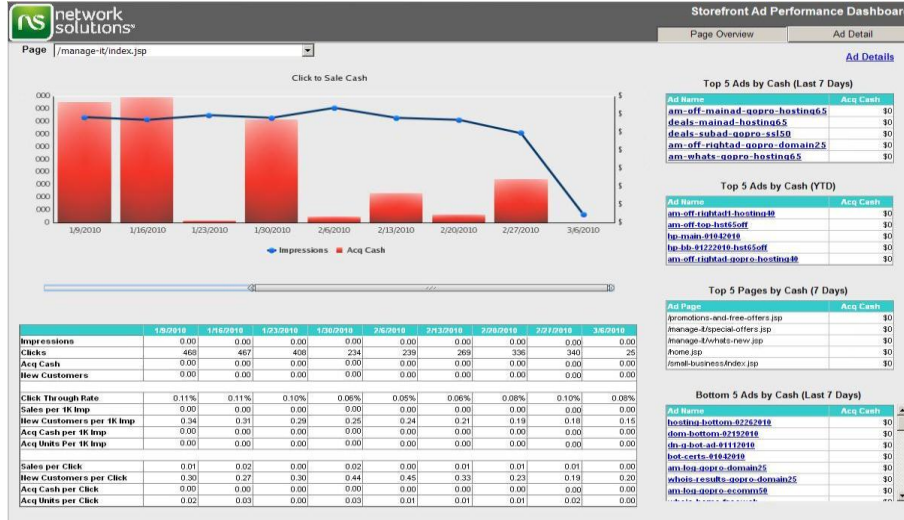

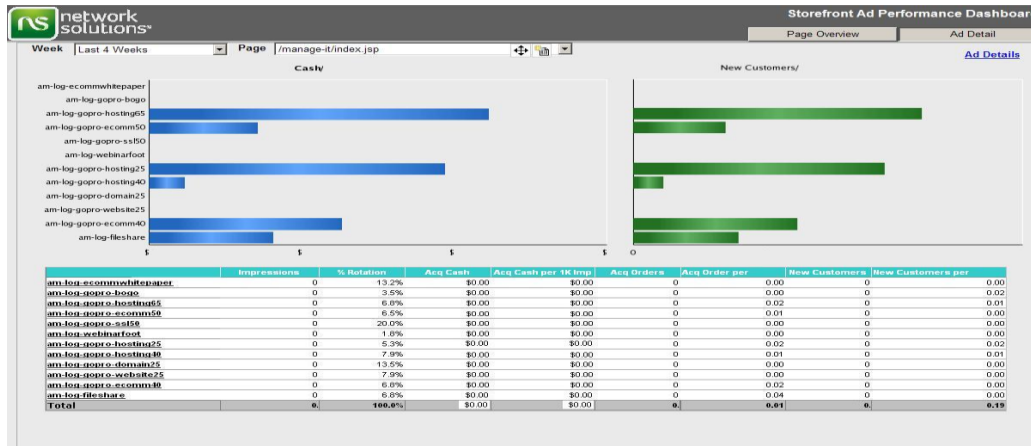**Figure 4 - SAPD Summary Dashboard View**

Figure 5 - SAPD Detail Dashboard View

Recommendation Engine Project

This section discusses the Recommendation Engine project in terms of project management and the IT infrastructure required to support the Recommendation Engine.

Project Management

The resources for the Recommendation Engine project filled the following roles at varying levels over the course of the project:

- Technical Lead
- Data Modeler
- ETL Developer

- Database Administrator
- Quality Assurance Engineer
- User Interface analyst

This effort did not include enhancements after the initial launch and development of the CMS system and its coordination with SRM.

IT development at Network Solutions uses the agile development method called Scrum.[6] All Scrum projects and product ideas are posted on a web portal for everyone to see. Each project has a weekly Scrum meeting of short duration, at which all contributors state what they have accomplished and what they plan to accomplish by the following week. Most of the Recommendation Engine was built using 30-day sprints (iterative cycles) between each new version of the software.

IT Infrastructure

To support the Recommendation Engine, the IT infrastructure consisted of the following components:

- Teradata and Oracle as the database management systems
- Informatica as the data integration tool
- MicroStrategy for reporting and analysis
- Salesforce to track sales activity
- Microsoft Sharepoint as a secure mechanism for sharing data and reports

Teradata Data Warehouse

The Teradata Data Warehouse collects, integrates, and distributes the shared data for many applications. It has evolved into the Enterprise Data Warehouse (EDW) for Network Solutions. As shown in Figure 6, the data warehouse consisted of a 3-node Teradata system, with one node provisioned as capacity on demand.[7] Specifically, the Teradata system is a model 5500H with 9TB of disk storage (of which 6TB usable for customer data), running the Teradata 12 operating system.
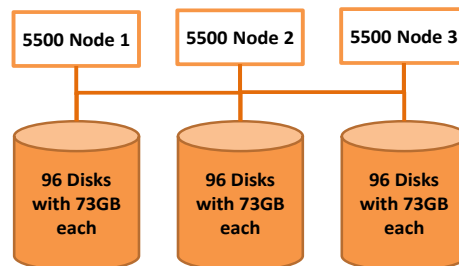
Figure 6 - Teradata Configuration

Figure 7 shows the various enterprise-level subject areas contained within the data warehouse. It is important to note that the Storefront project required data across many functional areas in order to perform its function.
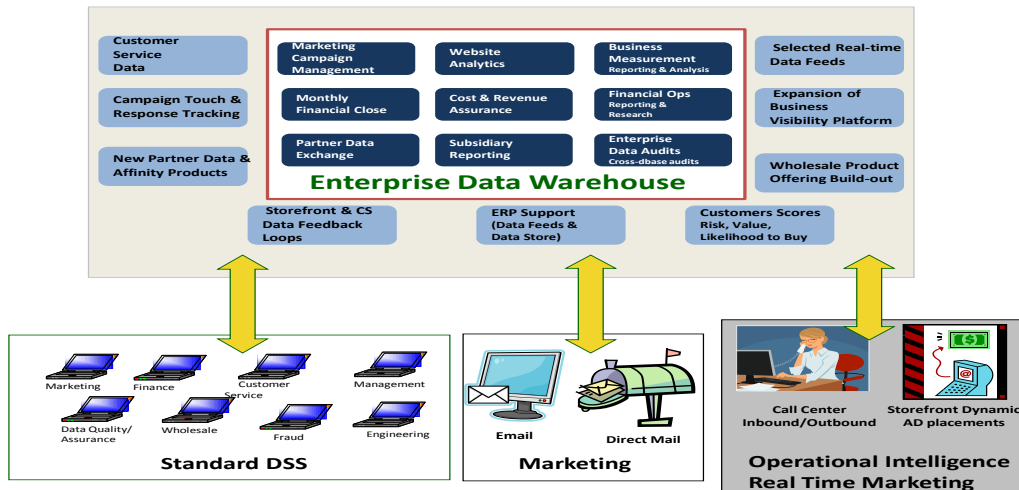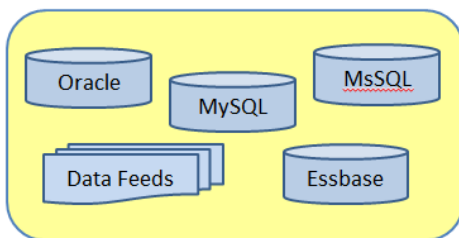
Figure 7 - Data Warehousing at Network Solutions

Data Acquisition

Figure 8 shows a simplified view of the data integration processes. A nightly ETL batch process acquires and loads data into the data warehouse using Informatica and Teradata FastLoad. In addition, a real-time data acquisition process data during the business day as needed, using Teradata TPUMP and web services.
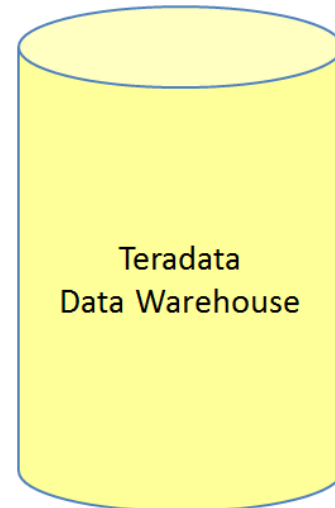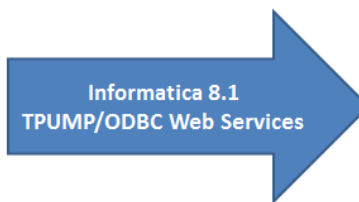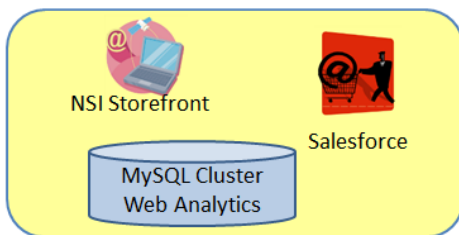
Figure 8 - Data Acquisition

Workload Management

One challenge of the Recommendation Engine project was balancing the diverse mixed workloads on the Teradata Data Warehouse to best utilize the computing and storage resources, which used the Teradata Active System Management tools.[8] Teradata batch jobs run overnight to process the previous day's data. It processes many different kinds of log data – access logs, commission engine data (supported by third party brokers/sellers), web click logs, etc.

The batch jobs include 5 primary workloads:

- Normal Batch, such as ETL processes;
- MicroStrategy Scheduled Reports;
- Ad hoc Queries;
- Tactical Batch, used by top management to troubleshoot exceptions from dashboards; and
- Tactical Queries, near real-time "Pulse" queries from sales persons while a customer is on the phone.

Benefits Realized

Network Solutions has been successful in using same Recommendation Engine concepts and mechanisms throughout the Storefront. All ads are placed on Storefront web pages using the Recommendation Engine, which has become pervasive across Network Solutions' entire infrastructure.

In summary, the business benefits realized from the Recommendation Engine project are:

- Automated process have eliminated manual analyses and reduced the cycle time of using the data.

- Personalized and relevant experience to our website visitors & existing customers by means of offers & educational advertisement, resulting to higher sale conversions.

- Better use of the real-estate on web pages for offer & educational advertisements of product lines

- An indirect impact was reduction in terms of cancellation calls to customer service for wrong products purchased.

- Accurate & audited data

- Flexibility & scalability in terms of on fly recommendation rule changes based on visiting patterns observed

- Storefront Advertisement Performance Dashboard was implemented using MicroStrategy to help track effectively how the recommendations-driven ads were performing. This enabled a direct insight that was actionable.

Additionally, the Recommendation Engine has assisted the call center in reducing handle time (for inbound calls) and research time (for outbound calls).

Lessons Learned

Each person interviewed was asked about lessons that they learned from the project and would share with other IT professionals. A number of insightful suggestions follow:

Innovative Flexibility

One of the lessons was the benefits from the in-house design and development team, which brought deep experience in the problem domain. Thus, Network Solutions can be very agile to business changes and reactive to new business requirements.

Prototyping is the Key

Creating prototypes as numerous stages of development and involving business users with reviewing those prototypes is critical to the project success.

Agile Development Approach

The Agile development methodology was an effective approach for Network Solutions. In particular, it helped to set clear and precise expectations of each development phase. Giving the business users the functionality expected accelerated adoption by building trust. Further, the more that was delivered, the more the business users requested.

Involve Experts in the Source Systems Used

The project team involved source system experts (both development engineers, database administrators, and quality engineers) with the Recommendation Engine project from beginning.

Balancing Session Data with Customer Data

Another challenge was to find the right balance of using session data (such as, session length) in combination of customer data (such as, products purchased). Network Solutions found that it is not one over the other. Both parts are needed to be combined properly to drive the recommendations. In other words, which attributes should be used to generate the most effective recommendations? Network Solutions starts with customer information and then enhances with session-related data, hence, re-enforcing the customer analytics. Based on the recent pages visited, Network Solutions may disqualify the top recommendation as inappropriate for this situation. In other words, session data should sometimes over-ride (or over-rule) customer analytics when making recommendations.

Balance ETL with Database Processing

When architecting a data warehouse, be sure to balance ETL processing with database processing for optimal performance and flexibility.

Audit against the System of Records

For any reports and analysis distributed, be able to audit against the System of Record. And, be prepared for questions about how these reports can be validated to show that they are correct.

Conclusions

The Recommendation Engine project has been deployed throughout Network Solutions' Storefront and pervasive across its infrastructure. It has also changed the way Network Solutions thinks about the interactions with its customers, moving from reactive to proactive.

The Recommendation Engine project required the integration of a diverse set of business data to be effective. To accomplish this integration, the Network Solutions team needed an extensive knowledge across cross-functional business processes. In addition, the IT infrastructure required the capability to store and managed this diverse data set, for which the Teradata data warehouse was a central element.

References

[1] Extracted from http://www.generalatlantic.com/

[2] Extracted from http://www.networksolutions.com/

[3] Statistical technique to estimate the response to a specific treatment, such as displaying a certain ad on a web page to a certain type of customers, See http://en.wikipedia.org/wiki/Propensity_score_matching for technical details.

[4] See http://www.exacttrend.com/weblogdb/

[5] Cookies are pieces of text from a website that the browser stores in cache. When the person visits the same website again, this cookie is sent to the website so that the person can be identified with their previous visit and are a convenience for remembering preferences and simplifying website interactions.

[6] Scrum is a rugby term used in an article by Takeuchi and Nonaka to describe a holistic approach to fast and flexible product development. It was adapted for software development methodology. For details, see http://en.wikipedia.org/wiki/Scrum_%28development%29. Note the discussion on this Wikipedia entry. There is more content there than in the actual entry and illustrates the confusion and disagreements within the Scrum community.

[7] Capacity on demand implies that this capacity is not charged by the vendor until it is needed by the customer.

[8] TASM is the Teradata Active System Management, which is a grouping of products to assist in the monitoring and management of the varying mixed workload of an active data warehouse. TASM consists of three products: Teradata Workload Analyzer, Teradata Dynamic Workload Manager and Teradata Manager. http://www.teradata.com/t/resources/brochures/Teradata-Active-System-Management-eb3625/?type=BR

About the Methodology

The objective of this case study is education—to share insights with other IT professionals so that we can mature as an industry, amid escalating business challenges and rapidly evolving technology. An on-site visit evolving a full day of interviews was conducted to adequately document the details of this case. Prior to the on-site visit, there were several telephone discussions that narrowed the scope to a specific IT project with its business requirements, timeline, resources, and results. As synthesized from the on-site interviews, several drafts were circulated for review. When the participants were satisfied with its contents, the document was submitted for publication approval by the company. All discussions and collected materials were considered confidential until the company approved the case study for publication.

Richard Hackathorn of Bolder Technology and **Dan Graham** of Teradata conducted the interviews. The persons interviewed at Network Solutions were:

- **Shiv Verma** - Director, Business Intelligence & Reporting

- **Navin Ganeshan** – Senior Director, Product Portfolio Strategy

- **Sathish Mallamula** - Sr. Manager, Business Intelligence Applications

- **Jagmeet Kaur** - Lead, Business Intelligence Operations & Infrastructure

- **Maribeth Knudsen** - Data Reporting Analyst

- **Ritesh Jain** - Manager, Business Intelligence Applications

About Bolder Technology

Bolder Technology Inc. is a twenty year old consultancy focused on Business Intelligence and Data Warehousing. The founder and president is Dr. Richard Hackathorn, who has over thirty years of experience in the Information Technology industry as a well-known industry analyst, technology innovator, and international educator. He has pioneered many innovations in database management, decision support, client-server computing, database connectivity, associative link analysis, data warehousing, and web farming.

Richard was a member of Codd & Date Associates and Database Associates, early pioneers in relational database management systems. In 1982, he founded MicroDecisionware Inc. (MDI), an early vendor of database connectivity products, growing the company to 180 employees and was acquired by Sybase, now part of SAP, in 1994. He has published numerous articles in DM Review and BeyeNETWORK. He is a member of the IBM Gold Consultants and the Boulder BI Brain Trust. He has written three books and was a professor at the Wharton School and the University of Colorado. He received his degrees from the California Institute of Technology and the University of California, Irvine.

About the Sponsor

Teradata is the world's largest company solely focused on creating enterprise agility through database software, enterprise data warehousing, data warehouse appliances, and analytics. Teradata provides the best database for analytics with the architectural flexibility to address any technology and business need for companies of all sizes. Supported by active technology for unmatched performance and scalability, Teradata's experienced professionals and analytic solutions empower leaders and innovators to create visibility, cutting through the complexities of business to make smarter, faster decisions. Simply put, Teradata solutions give companies the agility to outperform and outmaneuver for the competitive edge.