



# Lifting predictive analytics productivity at Sears

Paul Bibler and Doug Bryan

April 2005



*20 production models were migrated from the mainframe in 4 person-months.*

*By leveraging their data warehouse and adopting KXEN's new analysis technologies, Sears is doing more with less.*



Like many large retailers today, Sears faces fierce competition and an increasing pace of innovation. To meet these challenges it needs to cut costs while simultaneously increasing both productivity and agility. A case in point is Sears' catalog business. Three years ago the predictive analytics environment they used to model customer segments and target mailings was mainframe-based, difficult to interoperate with other databases, labor-intensive, and the specialized I.T. resources needed to support it were dwindling. The system was too expensive and inflexible. The team needed to find a way to be more productive and agile with fewer resources: a classic case of doing more with less.

They did so by implementing three approaches: (1) construct their own data mart within the corporate Tera-data warehouse, (2) automate many of the most labor-intensive steps in their analysis process using KXEN, and (3) automate the generation of model deployment code with KXEN so that models could be applied within the warehouse.

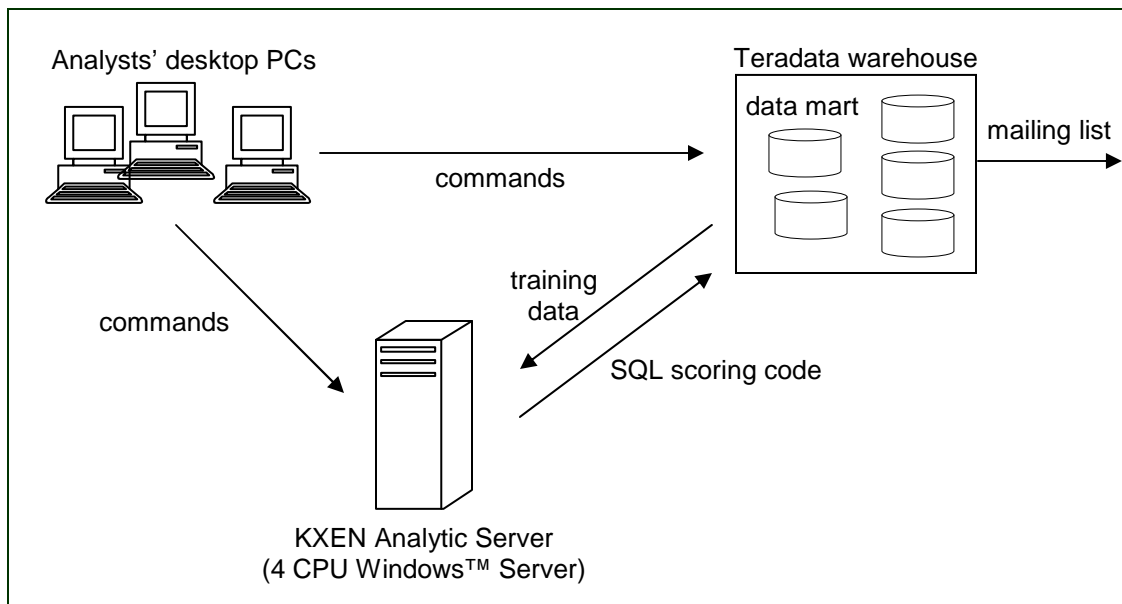
The first step was to build a data mart that combined data from Sears' multiple channels (stores, online, catalog), Sears brands (Sears, Orchard Supply Hardware, Lands' End, etc.), credit data, market demographics, and external sources such as mail-order overlays. The data mart includes over 900 attributes and is implemented within their corporate Teradata warehouse. Using the central warehouse gave the catalog team a highly scalable computing

and storage environment while reducing operational costs.

The second step was to automate some of the most labor-intensive, high maintenance, and repetitive steps in their predictive analysis process with KXEN's highly automated implementation of Structured Risk Minimization theory for data preparation, attribute importance, nominal attribute encoding, attribute value binning, combining bins, and selecting attributes without losing model accuracy or robustness.

The third step used KXEN to automatically generate model deployment code, known as scoring code, so that models could be applied within the data warehouse. In order to apply a predictive model to millions of customers, the computation of the model has to come together with the data of the customers. However the computation time is minimal compared to the time needed to move data from one system to another, so scoring within the warehouse means that they don't have to move the data and that saves a lot of time.

Sears' catalog team's new analysis environment is shown in the following diagram. A typical analysis proceeds as follows. Analysts prepare data in the data mart. This may include selecting particular customer segments, customers of certain product lines, or customers with purchases during a specific season. Next a small sample of the data, called training data, is copied to an analytics server and a model is trained. This data resides on the



analytics server for only a short period of time and is never maintained there, thus minimizing the operational costs of that server. Then scoring code for the model is generated in SQL by KXEN and loaded into the warehouse. Lastly the scoring code is executed in the warehouse on millions of customers to select the best ones for a particular mailing, and the resulting mailing list is exported to a commercial printer.

This system has allowed Sears to build more models and increase model quality while reducing development time and costs:

- Once the environment was in place, 20 production models were migrated from the mainframe in 4 person-months, resulting in hundreds of thousands of dollars in annual savings.

- They now train accurate, robust models in a few hours, and score 75 million customers in 30 minutes. In the old environment model training took weeks and scoring required several hours of very expensive CPU time.
- Automation has allowed analysts to spend their time where it can do the most good. They now build more sophisticated models than they could with conventional statistical analysis tools.

The cost, quality, and speed of modeling have all been improved by fully integrating with the Sears Data Warehouse, and by using KXEN for analysis. Increased analytic productivity also had an unexpected side-effect: business partners, analysts, and merchandising colleagues began asking more "what if" questions because they got responses so rapidly.

What if we remove that credit attribute? What if we add this seasonal data? Questions like these can now be answered in just 10 minutes. By leveraging their data warehouse and adopting KXEN's new analysis technologies, Sears is doing more with less.

**About the authors:** Paul Bibler (PBibler@Sears.com) is a business solutions consultant at Sears and has previous predictive analytics experience with Spiegel and Information Resources Inc. Doug Bryan (Doug.Bryan@KXEN.com) is a technical director at KXEN and has worked with emerging technologies at Amazon.com, SPSS, Accenture Technology Labs, and Stanford University.

---

KXEN provides next generation business analytics software that enables companies to use the customer information they already have to make better decisions. Based on a breakthrough mathematical theory, KXEN's products cut the time, cost, and staff required to generate accurate results. Visit us at [www.kxen.com](http://www.kxen.com)