

# CUSTOMER INFORMATION INFRASTRUCTURE

## 2005 COMPUTERWORLD HONORS CASE STUDY

### BUSINESS & RELATED SERVICES

A CUSTOMER INFORMATION INFRASTRUCTURE BASED ON AN INNOVATIVE GRID-COMPUTING INFRASTRUCTURE AND A SPECIALIZED GRID-ENABLED PROCESSING ARCHITECTURE ADDRESSES THE INCREASING INFORMATION MANAGEMENT AND TIME-TO-MARKET NEEDS OF THE COMPANY AND ITS CLIENTS, COST-EFFECTIVELY SCALES ITS OPERATIONS TO KEEP UP WITH DEMAND, AND MANAGES AN EXISTING INTERNAL "REFERENCE BASE" OF 20 BILLION RECORDS, AND TWO PETABYTES OF COMPANY AND CLIENT DATA. [20055353]



Robert Carrigan,  
Chairman of the Chairmen's Committee

Ron Milton,  
Vice-Chairman of the Chairmen's  
Committee

Dan Morrow,  
Chief Historian

### SUMMARY

At a time when no commercial products or methodology existed, Acxiom Corporation created a grid-computing infrastructure and a specialized grid-enabled processing architecture and strategy, Customer Information Infrastructure (CII), to meet the increasing information management and time-to-market needs of the company and its clients. The grid now consists of more than 4,000 nodes, and CII provides large-scale data management and content transformation processing for the creation and maintenance of complex business intelligence warehouses for Acxiom clients. With CII, Acxiom is providing greater value to its clients while cost-effectively scaling its operations to keep up with demand, currently managing an internal "reference base" of 20 billion records, and two petabytes of company and client data.

### APPLICATION

Since first stated in 1965, Moore's law has held true. The power of microprocessors, roughly correlated to the number of transistors, has doubled every two years. This has been good news for business and economies worldwide. However, as processing power increases, so too does the volume of data to process.

For many businesses, the keys to innovation and competitive advantage are locked inside this wealth of data – from internal sources on customers and prospects, to external sources of demographic and purchasing profiles, to supply chain data. But delivering high quality, timely, and actionable information and business intelligence requires more processing power than most organizations can afford to dedicate to the task.

Acxiom, a \$1.2-billion provider and integrator of business intelligence and hosted services-based solutions, found itself in the center of this data processing maelstrom. Acxiom was spending more and more money on traditional information technology, year after year, just to keep pace with client demand on its infrastructure.

Until 1995, Acxiom managed most of its data using IBM mainframe computers running the MVS operating system. As Acxiom grew, symmetric multiprocessing (SMP) technology became more powerful and cost effective. Acxiom transitioned its internal processes, such as data cleansing and data enhancement, and its clients' applications, to SMP platforms. While more cost effective than the mainframe, it was still expensive. To maintain its capability to manage the ever-growing volumes of data, Acxiom spent more than \$150 million a year for capital equipment.

To improve accuracy and breadth for marketing and other enterprise applications, Acxiom utilizes AbiliTec, a high-performance application the company developed to link and clean information on individual customers gleaned from multiple data sources. Acxiom matches every name and address it receives from clients against its 20 billion record in-house AbiliTec reference base, and assigns an identifier to each name and address record. Incoming client data is passed through a workflow process currently cleaning and standardizing addresses for more than 32 billion records per month. More than 40 billion records are "linked" per month, which equates to 15,000 links applied each second.

Processing speeds using SMP to support the AbiliTec process approached eight million records an hour. To meet the demand and growth of the business, about 20 such environments were needed to handle the processing load, which could have doubled Acxiom's computing expenses.

In 2000, Acxiom began looking for other options. Acxiom needed a way to get ahead of the data growth curve and off the information technology upgrade treadmill, and a platform to deliver a significant competitive advantage to its clients. As many of Acxiom's data processing procedures lend themselves to parallelization, a grid architecture looked promising.

In the spring of 2000, Acxiom initiated an R&D project to investigate using commodity PC-based servers in a grid to perform the building of its internal reference database and support the AbiliTec linking process. The performance improvements were staggering, averaging from a 10 to 40 times increase in throughput. Consequently, the company moved all of its internal services to the grid, eventually employing more than 4,000 rack-mounted, two-processor nodes. This number continues to grow.

In August 2003, Acxiom announced it would host client data and run client processes in that environment as well. That strategy evolved into a solutions architecture known as Customer Information Infrastructure (CII). CII is used to warehouse and provide quick access to all available data about customers and prospects for Acxiom's clients. CII also facilitates data cleaning, linking, data analysis, modeling, and other services for Acxiom clients at greater speed and capacity than otherwise possible.

## **BENEFITS**

Almost every business decision involves unwelcome trade-offs – one goal is achieved at the sacrifice of another. However, Acxiom's Customer Information Infrastructure and its grid computing environment eliminates many of the trade-offs faced by companies looking to improve marketing data quality and mine data for competitive advantage.

Before implementing its grid and CII, Acxiom faced trade-offs in speed, agility, rigor, or efficiency in the design of solutions for its clients. For example, a solution that takes two days to execute (the speed element) might take five months to change (agility). Also, that solution might require expensive technology to achieve that speed. For Acxiom's clients, CII addresses the technology aspect of the solution. It allows clients to use an arbitrary amount of data as part of their customer management solution, with application of an arbitrary amount of computing power to the processing of that data. CII removes the traditional trade-offs, serving as a facilitator rather than a dictator of data, speed, and storage capacities.

Acxiom client benefits associated with CII and grid processing include:

- Increased speed
- Smaller upfront capital expenditures for hardware
- Flexibility to add new applications or lines of business much faster and at lower cost

For Acxiom itself, scalability, reliability and operational efficiency are the main benefits of the grid and CII. Grid nodes are replicated across Acxiom's data centers for reliability and data protection. Acxiom worked with EMC to develop an online backup process, for a second level of data protection. The first grid application that went live in 2000 has never been out of service. From a scalability perspective, Acxiom can apply 300 million AbiliTec customer information links an hour compared to five to seven million links an hour with the previous SMP systems. The company now has the processing power to ensure that the most current data is available for use in its products and by its clients – a key factor in successful marketing campaigns and business intelligence applications. On the mainframe, it often took more than three months to update Acxiom's InfoBase database. On the grid, it now takes three days. That speed greatly improves the value of Acxiom's data products and provides Acxiom and its clients with a strong competitive advantage.

## **IMPORTANCE**

Acxiom's corporate strategy is to use technology to create a competitive advantage for its clients. CII is the latest and by far the most comprehensive of such initiatives.

Going against conventional wisdom, CII processing is not based on a commercial relational database, but on a specialized, highly parallel and redundant data grid for flat file storage, and a scalable workflow grid for customer data integration and client data-mart creation at high speeds. A relational database is used for a client's decision support functions using the intelligence refined from the flat files stored in the data grid.

More than 4,000 rack-mounted, two-processor nodes in Acxiom's data centers are dedicated to the grid to deliver the highest performance (as opposed to using idle time in computers elsewhere on the corporate network). Each node is a PC-based server running the Linux operating system. The grid servers are organized into "pods" that are logical groupings of servers housing the functions needed to handle various tasks such as data management, workflow, data transformation, analytics, and user interfaces.

The pods can be dedicated to specific clients or projects or used as a common resource for many clients or projects. The nodes in each pod operate together as one computer, managed by an Acxiom-developed grid operating system called Apiary Rex.

CII refers to more than just a virtual network of computers. It extends to the additional products, frameworks, and services developed to sit atop the hardware. This functionality includes:

- Mechanisms and processes for data management tailored to the grid;
- Mechanisms and standards for executing complex data transformations;
- A framework for transactional services;
- Relational database capabilities; and
- Delivery support for Acxiom products including data, customer data integration, postal products, suppression, and customer recognition.

Development of CII has resulted in a complete re-engineering of Acxiom's IT infrastructure and has totally reshaped the company's core IT principles. The results, and the benefits, are spreading well beyond the walls of Acxiom's data centers.

## **ORIGINALITY**

Though applications of grid computing have been around for a while, Acxiom has consistently remained two to three years ahead of the market in terms of deploying a grid for production purposes. While larger companies have struggled with the technology, Acxiom has succeeded by focusing not on building a general-purpose grid capability, but one focused squarely on the company's core value proposition. Acxiom's grid is optimized to perform the type of work required to deliver customer and information management solutions. Many of the standard data content (CDI & ETL) transforms that Acxiom performs to extract business intelligence from raw data are easily parallelized and run on the grid.

Acxiom's grid-computing initiatives have won awards and been featured in many articles. A few examples:

Information Week: True Grid

<http://www.Acxiom.com/default.aspx?ID=2605&DisplayID=18>

Acxiom Grid Innovation Wins 'Radical Data Warehousing' Award

[http://www.Acxiom.com/default.aspx?ID=2533&Country\\_Code=USA](http://www.Acxiom.com/default.aspx?ID=2533&Country_Code=USA)

Acxiom Pioneers Grid-Based Computing for BI

<http://www.gridtoday.com/04/0308/102802.html>

Acxiom Recognized as One of the Top 100 Agile Companies in CIO Magazine's 2004 CIO 100 Awards

[http://www.Acxiom.com/PrintVersion.aspx?ID=2565&Country\\_Code=USA](http://www.Acxiom.com/PrintVersion.aspx?ID=2565&Country_Code=USA)

## **SUCCESS**

Acxiom has transitioned all of its internal services and product builds to the grid. The success of Acxiom's AbiliTec customer data integration software (a core technology and differentiator), working against its 20-billion record reference base, would not have been possible or practical without the grid.

Currently 22 clients have migrated to CII. The goal is for every Acxiom client to take advantage of the infrastructure.

Following are some client success stories:

- Major Credit Bureau
  - Had 290 million records, representing 1 terabyte of data processed in CII

- Scored all records, resulting in 2 TB of data output
- Time to completion was 13 hours, including FTP time
- Large Credit Card Issuer
- Large file of 250 million customer records processed and scored in parallel using CII environment and mainframe
- Time to completion with CII was 15 hours versus more than 150 hours on mainframe
- Major Telecommunications Company
- With CII, the CDI processing time for 60 million records went from 5 days to less than 24 hours

In September 2004, Information Resources, Incorporated (IRI), selected Acxiom to provide data center operations and grid technology to support delivery of IRI's next-generation market insight and decision support solution – its “Content on Demand” factory. Using Acxiom's grid technology to process the volumes of data coming from daily retail, shipment, inventory, and consumer data updates, IRI is providing companies in consumer packaged goods (CPG) and retail with the ability to make decisions based on real-time information and gain a competitive edge.

“As part of our ongoing commitment to facilitate the reinvention of the CPG and retail industries... IRI is focused on developing new ways to harness information and provide profitable, actionable insights to our clients,” said IRI President and CEO Scott W. Klein. “Acxiom is a leader in information management innovations, offering the highly unique blend of expertise, operational excellence, and advanced data processing capabilities needed to support our aggressive initiatives and deliver increased value to the industry.”

“Acxiom has demonstrated that its information management platform along with its global data and services capabilities, when combined with IRI's content and analytics solutions, provide a powerful competitive advantage to the CPG and retail industries,” added IRI's CTO/CIO Marshall Gibbs.

## **DIFFICULTY**

Acxiom has deployed one of the largest, production grid environments in the world — if not the largest. When the project was started in 2000, not much of a roadmap existed. Acxiom built its own management functions when there was no commercial, generally available grid infrastructure resource management software. The Acxiom Products, Infrastructure, and Technology organization built a resource scheduler, grid control, and maintenance interfaces, software distribution functions, and grid-enabled data management functions. The company used open-source software, but where it was not available its associates wrote components themselves. During the initial build-out of the infrastructure, Acxiom consulted with major corporations developing general purpose grid products and confirmed that its technical direction was consistent with industry trends.