

## Enabling Online Analytics for Clients

SDI Increases Productivity, Responsiveness to Clients by Accelerating Massive Vertica Data Loads with Syncsort DMExpress

### Comprehensive Market Intelligence for the Healthcare Industry

SDI is a healthcare analytics organization that provides a range of market-intelligence services for the healthcare industry. It was founded in 1982 with the launch of a disease-surveillance program that tracked and projected illness and condition trends across the country. Today SDI gathers and presents data for a wide range of clients including all of the top 50 pharmaceutical/biopharmaceutical manufacturers. In addition, other top healthcare organizations, medical and surgical device manufacturers, consumer packaged goods retailers and manufacturers, government regulators, financial institutions and media groups rely on SDI for timely, accurate, and insightful data and analyses.

To serve such a diverse customer base, SDI aggregates massive volumes of physician-level and patient-level data derived from healthcare claims, hospital charges, electronic medical records, and more. The company delivers that data in various forms, along with a suite of analytical tools to access, analyze, manipulate and report.

**“What was taking us 18 hours through our old process now takes us only three hours.”**

— Chris McCaffrey,  
Managing Director of Data Warehousing

Collectively, that represents 25 TB of data in a massive Oracle data warehouse. SDI then creates dozens of data marts providing targeted subsets of data for different markets

and customers. Unfortunately, that had resulted in a series of performance-draining bottlenecks. “For many of our customer engagements, we have standardized on Vertica because, with pre-formatted data, it is very fast and enables our clients to perform online analytics, which is the preferred model today,” said Chris McCaffrey, Managing Director of Data Warehousing. “But the volume of data we needed to move and the lack of direct connectivity between Oracle and Vertica meant that we were experiencing delays during our monthly refreshes.”



#### INDUSTRY

- Business services — pharmaceutical analytics

#### CHALLENGE

- Improve speed and performance while loading data from 25 TB data warehouse into dozens of individual data marts
- Move data from Oracle to Vertica database
- Eliminate significant overhead of manually maintaining scripts and performing tactical operations

#### SOLUTION

- DMExpress data integration acceleration to improve extraction, transformation, and loading

#### BUSINESS VALUE

- Cut daily processing time from 18 hours to just three hours
- Improve staff productivity by minimizing maintenance overhead due to monitoring and restarting of jobs
- Increase ROI through automation, reducing overtime and manual process intervention
- Enhance customer satisfaction by accelerating on-demand online analytics on large data volumes

SDI's previous process called for taking a flat-file extract of data from the Oracle warehouse, compressing it, moving over the network, and loading it into Vertica. Those updates were taking as long as 18 hours for prescription and office-visit data marts.

"Although these were only monthly processes, they were problematic for our organization because they were processing-intensive and error-prone," said David Baker, Senior Database Administrator. "We were constantly running jobs over weekends, manually maintaining scripts (in Oracle PL/SQL), restarting failed jobs, and generally losing a lot of productivity to basic manual tasks."

### An Efficient Way to Load Data Marts

Aiming to streamline the way it loaded its various data marts — particularly the Vertica database — SDI consulted its technology partner, Blue Chip Technology, which recommended DMExpress from Syncsort. DMExpress is the only data integration software specifically designed to accelerate data performance and help organizations to significantly lower the total cost of owning and maintaining their data integration environments. DMExpress is used by hundreds of organizations to accelerate time-to-value, minimize maintenance overhead, and transform decision making through a highly optimized ETL engine.

"Most of us in the department were familiar with Syncsort from earlier in our careers," said McCaffrey. "We'd used it on the mainframe and it's a product that's well known for its reliability. Our warehouse is on an Unisys ES7000 Linux system with 96 processor cores and our data marts

run on HP ProLiant DL 380 servers, also running Linux. So we decided to put DMExpress into a proof of concept to demonstrate its fitness for our project."

According to Baker, that proof of concept was exceptionally easy. "Moving the data from Oracle to Vertica was very straightforward," he said. "There wasn't a lot to do. We selected some tables, defined some columns, and off it went."

The results were dramatically improved. "What was taking us 18 hours through our old process now takes us only three hours," said McCaffrey. "Depending on the month and the type of data, we end up moving as much as 90 GB in total. We did this to improve our internal IT infrastructure and the impact has been significant."

"It's been a nice change for me," Baker said. "I was spending a lot of time reloading jobs and babysitting the process. Now, we have a process that is more maintainable. The ROI we're seeing is in the dollars we're not spending on overtime for running manual processes."

"Could we have continued loading data marts the old way?" asked McCaffrey. "I suppose so. However, we would not have the speed, reliability, and maintainability that we enjoy with DMExpress and Vertica. Now that we have this in our toolbox, Syncsort has given us the power to be more responsive to our clients."

**"Moving the data from Oracle to Vertica was very straightforward. There wasn't a lot to do. We selected some tables, defined some columns, and off it went."**

— David Baker, Senior Database Administrator

## About Syncsort

Syncsort is a global software company that helps the world's most successful organizations rethink the economics of data. Syncsort provides extreme data performance and rapid time to value through easy to use data integration and data protection solutions. With over 12,000 deployments, Syncsort has transformed decision making and delivered more profitable results to thousands of customers worldwide.

**syncsort** RETHINK THE ECONOMICS OF DATA\*

50 Tice Boulevard, Woodcliff Lake, NJ 07677  
201.930.8200 | [www.syncsort.com](http://www.syncsort.com)