

Extreme Intelligence

Enabling In-depth Analysis of Claims Data at
The Centers for Medicare & Medicaid Services

The Centers for Medicare & Medicaid Services (CMS) has developed a process for providing ongoing, in-depth analysis on the tremendous amount of claims data received from each state. This enables CMS to determine trends, how money is spent, the public impact of certain diagnoses and drugs, and much more.

In order to handle the processing and analysis of all Medicaid claims data, CMS developed the Medicaid Statistical Information System (MSIS) project. According to Louis Gamerman, IT Specialist at CMS, the project provides a summary of data at the state level. Three data marts are used for querying: the State Summary data mart contains Medicaid information summarized by states; the Annual Person Summary data mart includes all of the Medicaid data summarized by individual; and the Clinical Services data mart includes drug- and diagnosis-related information. Prior to populating the data marts, raw data must be formatted and aggregated using a complex process.

“The reason we set up the data marts is because we’re not allowed to query the data warehouse,” says Gamerman. “It’s a giant, relational DB2 structure on the mainframe, and that structure is not indexed for querying. Also, because of the sheer volume of data involved, querying would bring the warehouse to its knees. Instead, we extract flat files from the underlying DB2 tables, where each flat file is equivalent to one table, and then convert them from EBCDIC to ASCII. For a state summary, we have six large files that we bring down with a combination of claims and beneficiary eligibility information, and we need to join them together and perform aggregates using complex business rules.” In order to complete the necessary processing, Gamerman turned to the consultants at ViPS – a leading provider of healthcare data management and analytics – to help develop the application.

“We turned to Syncsort because there was so much data, and we just needed the performance.”

— Bruce Crawford, Lead Systems Engineer, ViPS



INDUSTRY

- Government/Healthcare
- Headquartered in Baltimore, MD

CHALLENGES

- Providing analysis of claims data received from each state
- Inability to query the relational DB2 data warehouse
- Processing approximately 100 to 200 GB of year-to-date claims data per individual each quarter
- Accelerating joins and aggregates processing

SOLUTION

- Insert Syncsort DMExpress to seamlessly accelerate CMS’s Informatica environment
- With advanced sort, merge, and join functionality, DMExpress improves application performance and expedites analysis of very large volumes of claims data

BUSINESS VALUE

- Reduce total cost of owning and maintaining CMS’s data integration environment through more efficient use of CPU time, disk space, and memory
- Leverage existing Informatica investment for higher return
- Enable delivery of business intelligence reports with critical information, such as trend analysis, expenditures, drugs, and diagnosis, within the Medicaid Statistical Information System (MSIS) and Healthcare Information System (HCIS)

Drastically Improved Performance

According to Bruce Crawford, Lead Systems Engineer at ViPS, “The MSIS takes in a huge amount of data, and the CMS group had to calculate various sums and distinct counts across almost 1,500 permutations of the keys. It was a requirement of CMS to use Informatica to perform all of the business logic in the application. We turned to Syncsort because there was so much data, and we just needed the performance. We knew that we could count on Syncsort to complete the work and return it quickly. The product has been performing like a real workhorse, and we’ve also been taking advantage of its advanced merge and join functionality.”

The application now uses a combination of Syncsort and Informatica steps, along with custom C programs, to process approximately 100 to 200 GB of year-to-date claims data per individual each quarter. The flat files for each state and Puerto Rico are processed separately. The files are ordered and joined together using Syncsort, while Informatica is used to apply the business rules and perform the aggregation. “The performance is drastically better using Syncsort prior to performing the complex aggregations. We also do a lot of passes and a lot of steps between Syncsort and Informatica. Applying the business rules can get complex, and then we’ve got to do other tasks, such as lookups – plus there are dimension tables involved, so we’re doing surrogate key generation. And because of the volumes of data, Syncsort provided us with the only sensible way to join the files,” Gamerman added.

For the MSIS, the data is processed on a quarterly basis using a Sun E10000 running Solaris. The aggregates are then loaded into the three data marts and used to populate IBM Cognos cubes for querying. The information can be used to create reports that show a state’s Medicaid expenditures by type of service and age group, Medicaid beneficiaries by type of service and age group, Medicaid expenditures for programs broken down by maintenance assistance status and basis of eligibility, and much more.

Leveraging Every Feature

CMS also decided to integrate Syncsort into its Healthcare Information System (HCIS) project due to its similar requirements for high-volume number crunching and summarization. The HCIS handles the processing and analysis of all Medicare claims. As with the MSIS, the claims-level data in the HCIS is migrated from the mainframe to the UNIX server, then a series of Syncsort and Informatica steps are completed. “We compress the data, performing approximately 200 permutations of the data, and then load it into Oracle tables,” says Crawford. “We complete the entire process with Syncsort and Informatica. We also use all of Syncsort’s features to do our joins, merges, sorts, and more.”

Gamerman concludes, “The primary reason we chose Syncsort for both the MSIS and HCIS projects was because we needed a really powerful tool that could manage the volumes of data that we were talking about. Syncsort was the solution that could handle the sheer volumes of data and claims we had to join together. It has also made efficient use of disk space and memory.”

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