

# Clickstream Data Integration

Delivering the fastest and most scalable technology to support clickstream data warehouse initiatives

## SYNCSORT PROVIDES OPTIMAL SUPPORT FOR SUCCESSFUL CLICKSTREAM DATA WAREHOUSE INITIATIVES BY:

- Quickly processing massive amounts of clickstream source data and easily meeting overnight batch windows for Clickstream Data Warehouse (CDW) refreshes
- Eliminating the need for additional, costly hardware investments by dramatically reducing CPU time, elapsed time, and disk I/O – all while running on commodity hardware
- Deploying a graphical, drag and drop development environment, resulting in implementation in days, not months, ease of application development and maintenance, and a 75% increase in productivity over hand-coding

## The DMExpress Clickstream Data Integration Solution Enables You To:

- Gain faster access to actionable web analytics information, allowing stakeholders to make better, more insightful decisions
- Identify previously unseen data trends through processing granular data over significant time periods
- Provide a scalable solution to accommodate growth of website data volumes
- Enhance developer resources by enabling individuals of wider skill levels, to easily develop new clickstream data transformations

**“When we implemented DMExpress, comScore saw a 500 percent improvement in clickstream data integration throughput, providing us with a sizeable competitive advantage.”**

— Michael Brown, comScore Inc.

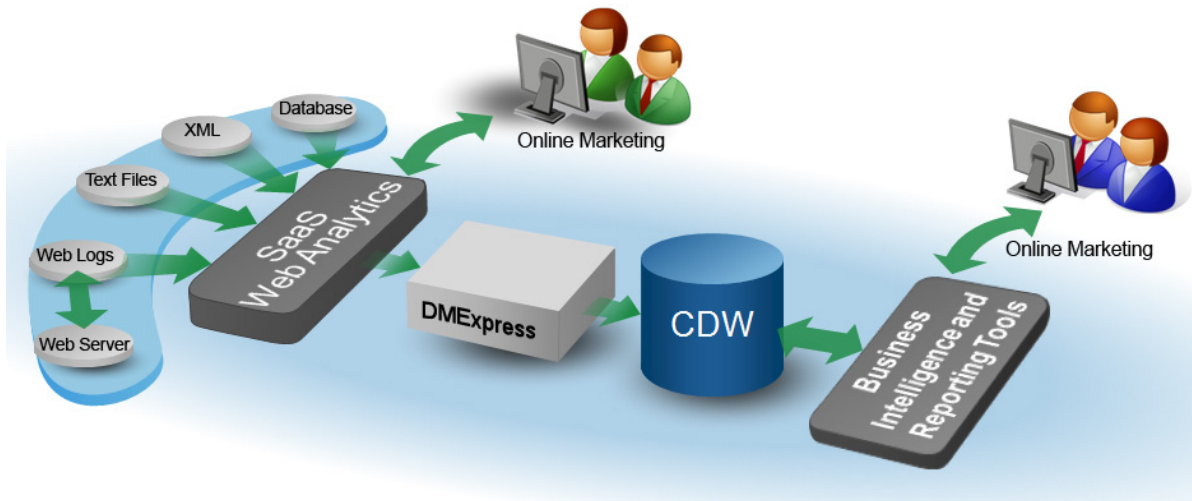
## Business Challenge

Access to timely clickstream data from a company’s website provides insight into online visitor behavior and search patterns, which in turn enables companies to be more effective in myriad ways, including improved pricing, more effective campaigns and offers, better visitor segmentation and targeting, and optimized website layout and workflow.

As clickstream data grows exponentially and companies are becoming more sophisticated in their web analytics requirements, the need to augment off-the-shelf web analytics applications has become critical. Many organizations are turning to building and maintaining an internal CDW for advanced segmenting, analyzing, and reporting critical to decision-making.

CDWs have been difficult to create for many companies because of the volume and complexity of the data involved – both the raw source data files that need to be cleansed and integrated, as well as the resulting warehoused data to be analyzed. The raw clickstream source data, which often exists as huge, complex text files, must be parsed, structured, cleansed and loaded on a regular – sometimes daily – basis into the CDW before the CDW can provide value. This integration process often introduces additional cost, complexity, and delays.

In addition, companies frequently turn to creating their own custom-coded solutions, which are prone to risks and errors, and difficult to maintain or extend.



## Solution

Due to the massive data volumes and the complexity of clickstream data, a successful CDW initiative will require high-performance data warehousing and data integration technologies. High-performance data integration technology is required to initially populate the CDW, and to refresh the CDW on an ongoing, daily basis.

Syncsort's Clickstream Data Integration (DI) Solution is optimized for processing massive amounts of clickstream source data at record speed, while leveraging commodity hardware. The complete solution consists of DMExpress software, professional implementation services, and training modules specifically tailored to implement techniques and patterns optimized to streamline the data integration process of CDW initiatives.

### DMExpress Software

DMExpress, Syncsort's high-performance data integration software, eliminates data latency by significantly reducing CPU cycle time, data transformations elapsed time, and disk I/O activity while utilizing minimal resources on commodity hardware. It also shortens deployment and development timeframes by leveraging a graphical, drag and drop development environment.

With the Clickstream DI Solution, DMExpress provides out-of-the-box features that include sophisticated and flexible text/string processing, "URL decode" functions, and extensive pattern. DMExpress also provides the ability

to quickly process raw web server log files, as well as numerous web analytics vendors' data formats/files.

### Professional Services, Training, and Best Practices

Syncsort's Professional Services teams are skilled in working in numerous web analytics production operating environments. They offer services and training specifically tailored for the optimal execution and management of data integration processes critical to CDW initiatives.

### Customer Success Story

comScore, a global digital media provider, collects terabytes of clickstream data from over 2 million web users. They were experiencing challenges processing and integrating over 40 billion new master records per month. Hand-coded solutions were initially attempted to solve the problem but could not handle the performance needs.

comScore needed a data integration solution that would deftly transform terabytes of raw clickstream data into useful, granular information at high speed, and rapidly load it into a Sybase IQ Analytics Server. They evaluated various solutions and chose DMExpress because of its extreme data integration performance and high return-on-investment. With DMExpress, comScore is now able to process billions of records 5 times faster than with its previous hand-coded programs, using off-the-shelf Windows and Linux machines.

## 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.