

# DMExpress Extreme Data Integration

The fastest & most resource-efficient way to extract value from data

## Navigating in the World of Big Data

Your organization needs to make the fastest, most accurate, and most insightful decisions in increasingly complex business situations. Yet growing data volumes, shrinking operational windows, and demands for immediate information are all converging like a perfect storm.

- Are these dynamics putting pressure on your ability to deliver dependable, actionable business intelligence?
- Are you concerned that this perfect storm may cause severe bottom-line ramifications to your business, including lost revenue opportunities, increased costs, impaired decision making, and customer attrition?
- Are you in need of a faster, more resource efficient, more scalable way to unlock the business value of your data just to keep up?

Though your information technology architectures are straining under the pressures of this storm, odds are you do not have the luxury to “rip and replace” your existing environment with new technologies optimized for the world of big data. You need to extract more value from your existing investments by finding the means to efficiently transform, integrate, and move data within required timeframes, even as your data volumes explode.

## DMExpress Delivers Extreme Data Integration

To help weather this perfect storm, Syncsort’s extreme performance data integration software, DMExpress, empowers your organization to rapidly and effectively deliver information from massive and complex data. With DMExpress, your organization eliminates lost revenue caused by processing bottlenecks, while using significantly less CPU and memory resources, dramatically reducing your hardware investment.

DMExpress, which can be deployed standalone or complementary to an existing data integration platform, is non-disruptive to your existing environment, is cost-effective, and is simple to implement and maintain.

### DMExpress enables your organization to:

- **Achieve extreme performance for data integration processing**
- **Unlock previously unattainable value to the bottom line**
- **Transform decision making**

### EXPERIENCE EXTREME

- DMExpress is the fastest data integration software at 1.6GB/second
- DMExpress delivers cost-effective scalability with a minimum resource footprint
- DMExpress enables ROI in less than 6 months

## Achieve Extreme Performance for Data Integration Processing

DMExpress is the fastest and most resource efficient data integration technology available. It delivers highly efficient data integration solutions by processing massive data sets in minimal elapsed time, and provides cost-effective high scalability by leveraging a minimum resource footprint on commodity hardware. Superior performance differentiation is achieved through proprietary transformation algorithm design, dynamic optimization of those algorithms at run-time, and exploitation of the underlying system resources DMExpress is running on.

These three primary capabilities are what makes DMExpress performance stand above the rest:

**Algorithm Design.** Compared to competitors, DMExpress utilizes a vast library of transformation routines and proprietary algorithms for functions such as sorting, aggregating, joining, parsing and pattern-matching.

**Dynamic Optimization.** DMExpress is optimized to programmatically select the most efficient algorithms based on the data structures and system attributes it encounters at run-time. The data integration application behavior adapts and optimizes itself to the exact characteristics of any particular job and system.

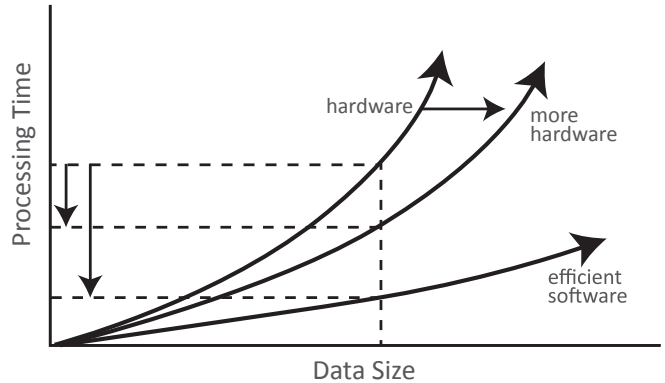
### DMEXPRESS SUCCESS STORY – COMSCORE

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.

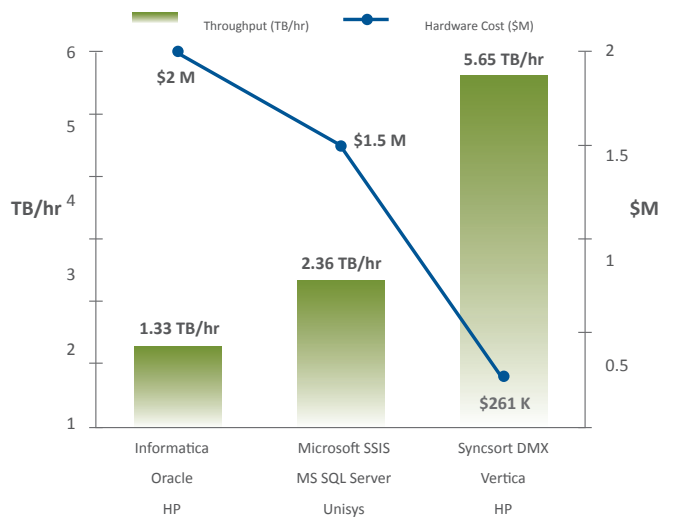
comScore needed a data integration solution that would deftly transform terabytes of raw clickstream data into useful, granular information at high speed, and rapidly load it into a Sybase IQ Analytics Server. They evaluated various solutions and chose DMExpress because of its extreme data integration 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.

**Architecture Exploitation.** DMExpress is calibrated to perform optimally for any specific hardware and system architectures, taking into consideration parallel processing capabilities, I/O speeds, number of processors, instruction / memory cache resources and others.

**Figure 1. Efficient algorithms perform and scale better than adding more hardware resources**



**Figure 2. DMExpress Sets World Record: Delivering the Fastest Data Integration Speed**



DMExpress extreme performance data integration software has been proven to be the fastest and most resource efficient way to deliver information from massive and complex data. DMExpress is the only data integration solution to break the 1 GB/second barrier. In an independently audited benchmark, DMExpress extracted, transformed, and loaded 5.65 TB of raw TPC-H data at a rate of 1.6 GB/sec. The feat was accomplished on low-cost commodity servers, setting a dramatically lower cost standard for data integration throughput.

## Unlock Previously Unattainable Value to the Bottom Line

As organizations rethink the economics of their data integration strategies, they must consider technologies that unlock greater levels of business value from their data than had previously been achieved. This enhanced value can come in many forms: new and increased revenue streams, reduced operating costs, rapid time-to-value, and extracting more value from existing infrastructure investments. By delivering extreme data integration, DMExpress helps organizations achieve these bottom line benefits.

Actual examples of revenue generating value enabled by DMExpress for Syncsort customers include:

- Executing 4 times more database marketing campaigns every month
- Enabling the creation of new data-driven customized applications for clients
- Offering more timely and targeted up-sell promotions to existing insurance policy holders
- Delivering new information-based services
- Cost savings and higher ROIs have been achieved through:
  - Accelerating performance in Informatica, IBM DataStage, Ab Initio, and other data integration platform environments
  - Delivering World Record economics at a processing cost 90% less than for other major products
- Enabling rapid deployment, with a 66% average reduction in deployment time
- Eliminating costs associated with custom coding

### DMEXPRESS SUCCESS STORY – OVERSTOCK.COM

Overstock.com, a major online retailer, relies on clickstream data from its web site for analysis by its brand and category managers. To achieve granular and customer-centric views, they maintain an internal clickstream data warehouse. To parse, structure and cleanse the clickstream data files, they had used complex shell scripts which proved difficult to maintain.

Overstock learned that DMExpress offered an agile development environment, quick time to deployment, fast and efficient processing of file-based data, rapid delivery of information, and scalability. By replacing their custom-coded scripts with DMExpress, Overstock reduced costs by decreasing processing time by 65% and enabling more efficient application development. DMExpress also enabled revenue growth through its inherent extensibility, allowing Overstock to easily bring on new suppliers of inventory to generate new revenue streams.

**Figure 3. Typical architecture of a clickstream data warehouse implementation with DMExpress**



## Transform Decision Making

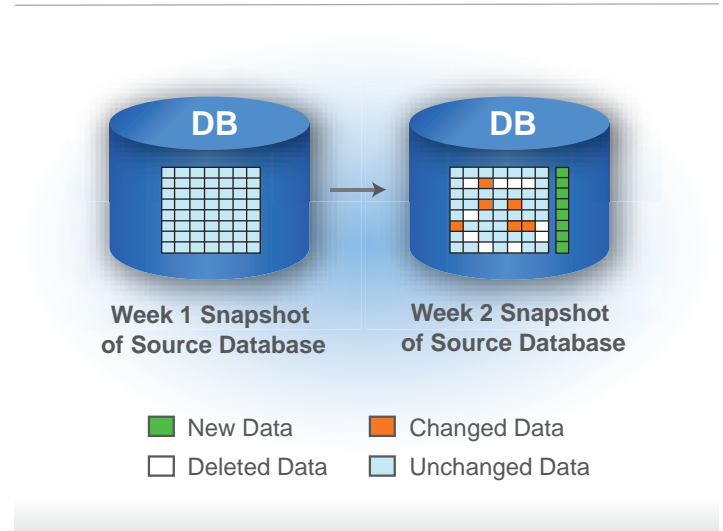
Prevalent in today's business intelligence environment is the need to support day-to-day decision making via operational data stores. Examples include customer call centers that require the latest customer information, and inventory forecast models which need current stocking level detail to optimize supply chain information. This need for current, low-latency operational data relies heavily on the performance of the associated data integration technology. An operational data warehouse, for example, must be refreshed with current data as often as needed, and lengthy jobs to integrate new information are not acceptable for meeting low-latency demand.

The extreme performance of DMExpress makes it the ideal integration technology for operational data warehousing, and it empowers customers to make better, faster and more insightful decisions in increasingly complex business operations.

An example of a key DMExpress performance capability that is crucial for operational data integration is change data capture (CDC), or delta processing. Using CDC, an older version of a data source is compared with the most current one and only the data records that have changed

are identified, extracted, transformed, and loaded into the warehouse. By eliminating the need to do a complete and protracted data load every time new data becomes available, processing is accelerated and new intelligence is available to decision makers sooner.

**Figure 3. Changed Data Capture Accelerates Data Warehouse Loading**



### DMEXPRESS SUCCESS STORY – PINK OTC MARKETS

Pink OTC is the third largest U.S. equity trading marketplace. They needed to improve their reporting capabilities to stay ahead of the millions of records their trading systems amass daily.

Integrating into a Vertica Analytic Database, DMExpress slashed the time it took to generate market reports from hours to seconds, and increased customer satisfaction. By filtering, analyzing, and aggregating historical and streaming financial data into easily accessible reports, Pink OTC now provides its customers with actionable intelligence they can turn into opportunity.

## Chosen by Thousands of Customers

For over four decades, more than 6,000 customers, including 90% of the Fortune 100 have used Syncsort's data integration technologies and services to deliver previously unattainable business value by integrating more data in less time using fewer system resources. Driven by the need to remain competitive, companies in numerous industries have used these technologies to solve a vast number of business problems.

DMExpress is used in industries as diverse as:

- Financial Services
- Retail
- Manufacturing
- Technology
- E-Commerce
- Telecom
- Healthcare
- Media
- Education
- Government
- Insurance
- Communications

### DMEXPRESS EXTREME PERFORMANCE BUSINESS SOLUTIONS

*Present-day business solutions for which DMExpress is commonly used include:*

**Data Integration Acceleration** – The supporting technologies and services of the DMExpress Data Integration Acceleration Solution augment and complement existing data integration environments to effectively eliminate specific performance bottlenecks. To achieve this, DMExpress can easily be invoked into data integration platforms from Informatica, IBM, Microsoft, and many others.

**Clickstream Data Integration** – Integrating data into a clickstream data warehouse requires massive speed and scale, complex text processing, and a flexible development environment. The DMExpress Clickstream Data Integration Solution provides these capabilities and can be used with raw web log data or web analytics tools.

**Mainframe Application Modernization** – The DMExpress Mainframe Application Modernization Solution enables successful migrations of legacy mainframe applications to open systems. It translates and transforms large amounts of data quickly, even data with complex formats, and supports various migration approaches including re-hosting.

**Data Integration for Database Marketing** – Effective database marketing campaigns require detailed, accurate, and current customer information, but integrating this data can be difficult due to large and disparate data sources. The DMExpress Data Integration Solution for Database Marketing addresses these challenges because it handles massive speed and scale, support for complex conditional logic, and easy development.

## DMExpress Quick Facts

Key Features	
High-Performance Data Transformations	Transformations include high-performance aggregations critical for query processing, high-performance joins to speed lookups, advanced text processing, and source/record/field level operations.
Rapid Development and Deployment	Development and Deployment features include job flow mapping, custom task piping, responding to rapidly changing application environments or constraints, and Expression Builder for defining data transformations (define values and conditions, which are frequently an implementation of business rules).
Management and Analysis	Tools include Job Editor (task assembly, integration with third-party applications, script generation, scheduling, execution, and monitoring), impact and lineage analysis for identifying the effect of input variations on outputs, and source/target data descriptions.
Ease of Use	Features include metadata layout export for simplifying metadata exchange with other data integration tools and event-based scheduling (no hand coded scripts required).
Efficient Resource Utilization	Efficient resource use is achieved through architecture exploitation, data compression of flat files to use less disk space and speed performance through reduced I/O time, advanced checkpoint/restart to only rerun tasks not already complete, and others.
Technical Specifications	
Data Sources and Targets	DMExpress provides connectivity to the following sources and targets: Oracle, SQL Server, DB2, Sybase, MySQL, Teradata (including pushdown optimization and load customization), Netezza, Vertica, Microsoft Access, flat files, XML files, and others. Additional sources supported include Mainframe and SAP sources.
Operating Systems Supported	Include Windows (on Intel and AMD), UNIX (AIX, HP-UX, Solaris, Tru64), Linux (Intel, AMD, IBM PowerPC, IBM System z).
DMExpress Professional Services and Training	
24x7 Customer Support	Staffed by experienced engineers.
DMExpress Product Training	Utilizing Syncsort's data integration training experts, your team will learn to perfect its data integration processes in only a few days.
Fast Start Services	Help you build your Data Integration foundation.
Performance Assessments	Syncsort engineers analyze your environment to assure optimal configuration and best practices for maximum performance.
Customized Services	Robust personalized assistance maximizes the benefit of your DMExpress investment for years to come.

## 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<sup>®</sup>

50 Tice Boulevard, Woodcliff Lake, NJ 07677

201.930.8200 | [www.syncsort.com](http://www.syncsort.com)