

white paper

## expressor/Netezza intelligent load & go<sup>tm</sup>

By Dave Shuttleworth, principal consultant and co-founder,  
Edge Associates and John Traves, principal consultant  
and co-founder, Emunio Consulting

March 2009

As the volume of data expands and the need to analyze it follows, most organizations find themselves in one of two camps: using tools to extract, transform and load data into new target databases for analysis (ETL), or extracting and loading data into new targets and performing the transformations in those databases (ELT).

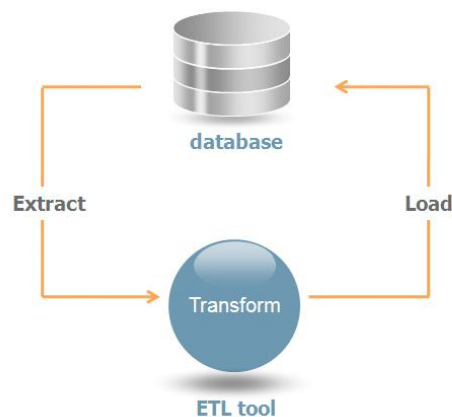
Each approach is best-suited to solving specific problems, but cost barriers and vendor lock-in have forced most organizations into one camp or the other and limited their ability to move between the two.

The expressor/Netezza intelligent load & go™ solution breaks down these barriers and allows organizations to choose the data warehouse architecture – ETL, ELT, or a hybrid - that best fits their business needs.

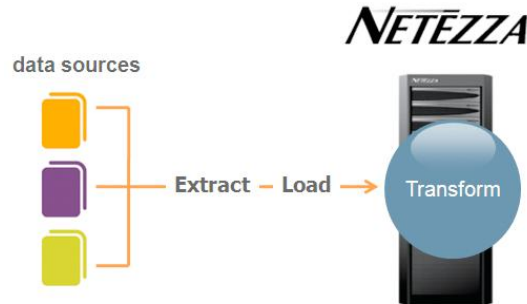
### ETL vs. ELT: high cost of ETL tools limit their adoption

Two seemingly unrelated events created an ETL/ELT divide in the data integration community. The first was the introduction of data integration tools, which were simple code generators meant to be used with the flat-file databases of the time. The second was the introduction of relational databases. Both had limited functionality initially, but have evolved over time.

ETL tools expanded to address enterprise data integration needs, and the top-tier ETL tools evolved into data integration suites, offering sophisticated features and functionality, such as metadata repositories, data lineage, profiling, governance, impact analysis and configuration management.



While slightly slower to evolve, relational databases more than compensated for their original deficiencies, and the latest generation of products, such as Netezza’s massively parallel **appliances**, perform many data integration tasks very effectively, especially if the data continues to reside in the same database – the classic ELT scenario.



As a result, both ETL and ELT are now capable of transposing vast volumes of data but the strengths of each approach mirror the weaknesses of the other.

“Just as Netezza revolutionized the data warehouse market with its high-performance, low-cost appliance, expressor software’s semantic data integration system is making it possible for more organizations to afford high-performance ETL and DI solutions. By deploying expressor for intelligent load & go, Netezza customers can choose the application architecture that fits their business needs.”

- Rick Barton, managing director and co-founder of Emunio Consulting

Today’s graphical ETL tools offer high throughput and faster development. But aspects of the ETL design and specification process, such as determining the rules for data matching, merging and cleansing, can consume a disproportionate share of the effort in any project. And high ETL licensing and maintenance costs have been a serious barrier to widespread adoption.

Compared to ETL, ELT offers lower costs and a more simplified architecture, as the database becomes a single platform for all functions and SQL is a common and familiar coding environment. On the other hand, ELT makes it difficult to capture and trace valuable metadata in process flow and business logic.

The result is that companies are often locked into ETL or ELT for architectural, cost or performance reasons, but either choice can be a compromise.

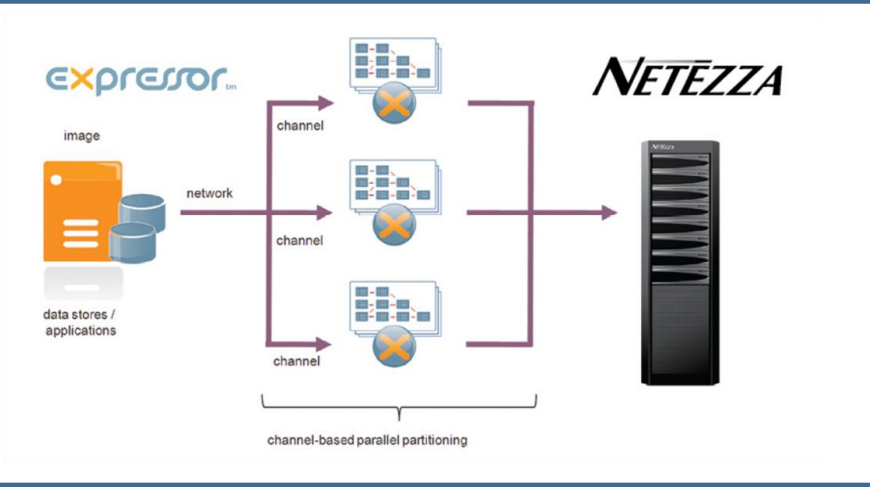
### **expressor/Netezza: an affordable intelligent load & go solution**

Users of the Netezza data warehouse appliance know the benefits of performance, simplicity, scalability and cost-effectiveness gained from a new product that builds on past experience with the latest technologies. In the ETL/data integration arena, expressor

software has taken a similar approach, offering unique semantic integration capabilities in a next-generation system for effective implementation of data warehouses and marts.

The [expressor semantic data integration system](#)<sup>tm</sup> is a disruptive technology that changes the metrics of performance, functionality and price through its ultrafast parallel data processing engine, semantic rationalization process and affordable, usage-based pricing model.

The combination of expressor and Netezza in intelligent load & go offers the benefits of both ETL and ELT at an affordable price – and allows organizations to choose the data warehouse architecture most appropriate for their specific business objectives and application demands.



Intelligent load & go leverages expressor’s collaborative, [role-based approach](#) to encourage best practices in development and data governance by allowing developers, architects, data stewards and analysts to share a common semantic metadata repository. expressor’s unique [semantic rationalization](#) process promotes greater reuse of business rules and improves organizational understanding of the relationships between physical and business metadata.

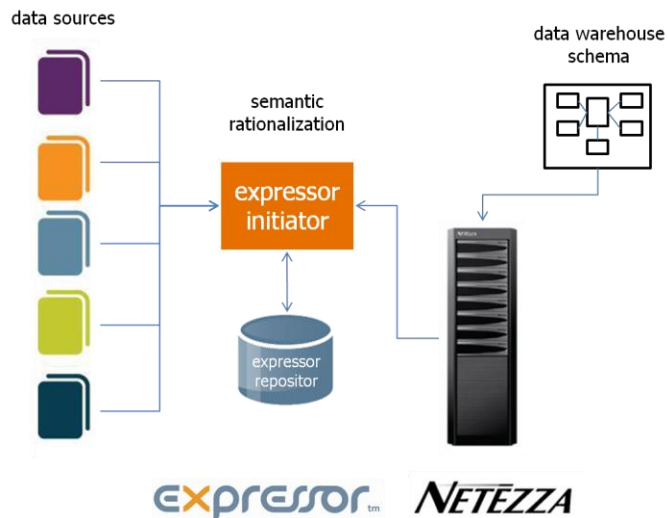
For maximum data throughput, expressor includes an “nzload motor” that integrates seamlessly with the Netezza nzload bulk loader. Multiple parallel load streams are easily configured (even to the same table) and data can be “streamed” via expressor from a source database or file system (across a network if necessary) directly into a Netezza database via nzload.



Since expressor is capable of extracting metadata from tables in the Netezza system catalog and performing semantic rationalization on column names, intelligent load & go provides a sound and consistent foundation for incremental development, fast deployment and rapid return on investment.

### deployment scenarios

To jumpstart an initial data warehouse with expressor/Netezza intelligent load & go, an organization would simply load the desired data warehouse schema into Netezza, rapidly build “common business definitions” from the Netezza system catalog via expressor’s initiator (a powerful metadata bulk loader), and then semantically rationalize the source metadata with these common business definitions. All metadata are stored in the expressor repositor, an enterprise-class semantic metadata repository that collects, stores, and manages project management information, reusable data descriptions, application file versioning, performance metrics, and the implementation and enforcement of role-based security.



Once the source and target metadata are semantically rationalized, expressor provides easy-to-use, role-based GUI tools for building expressor drawings, which are graphical data flow charts describing the movement of data from various sources to the Netezza based data warehouse and the operations performed on the data as it is being transformed.



In addition to expressor [illustrator](#), a Microsoft Visio-based developer application for creating these drawings, and the above-mentioned expressor initiator, the expressor system includes a Microsoft Excel-based application through which business and data analysts define reusable business rules based on the common business definitions. The expressor system also includes a Web-based administration tool used by data architects for creating “images” and “networks,” which are the data and network descriptions used by the application developer. These and other expressor tools provide a complete set of collaborative GUI tools for developing comprehensive data integration applications and managing these applications throughout the project lifecycle.

The benefits of intelligent load & go increase exponentially as an organization develops additional applications, since the more you use expressor, the more it learns about your business and the semantic rationalization process becomes increasingly automated. This increased “intelligence” leads to greater re-use, shorter development times, improved data governance and facilitates the addition of new data sources.

Developed expressor applications can be easily deployed using expressor’s high-performance parallel data processing [engine](#), which has been designed to be as compute- and memory-efficient as possible and take full advantage of modern processor architectures including 64-bit processors, multi-core CPUs, large SMPs and even MPP systems, run in batch or real-time and process virtually any type of data.

When tested in a Netezza environment (dual core Windows XP machine with 2GB RAM and a gigabit connection to an 8-rack Netezza NPS) the expressor engine delivered a sustained throughput of 35.5MB/second on a single expressor [channel](#) (unit of processing). Using four channels would saturate the NPS at 500GB/hour. expressor has committed to scaling this performance and maintaining tight integration with all new versions of the Netezza appliance.

## best practices

To optimize the performance of your expressor/Netezza intelligent load & go implementation, we recommend the following best practices:

- Define best practices early: For Netezza, optimize data types for space utilization, use bulk INSERT/DELETE for maximum performance, watch stats, zone maps, materialized views, and TRUNCATE vs. DELETE to optimize tuning, and pay attention to backups, reclaims, statistics collection and other operational functions. For expressor, make sure to follow standards for coding and file naming.
- Encourage parameterizations and reusability within procedures: expressor encourages large-scale reuse through semantic rationalization, but also look for opportunities to build parameterized or metadata driven applications to cut down on rework. expressor provides multiple levels of configuration files and variables to enable parameterization. Many users load staged files via parameter-driven code.
- Identify and allocate technical roles: even if one person may have multiple roles at the start of a project, documenting everything at the outset enables smoother transitions once a project expands or when personnel inevitably changes
- Start the semantic rationalization upfront by connecting expressor and Netezza early: The benefits of this approach are illustrated in the “greenfield” deployment scenario.
- Understand and embrace parallel processing: For Netezza, make sure to optimize data distribution. For expressor, training and documentation will enable the user understand the behavior of data when it is split into parallel streams.
- Perform granular transformations in expressor to benefit from metadata functions; aggregate in Netezza to benefit from performance: This configuration plays to the strengths of each technology, but the choice of where and when to aggregate and transform will ultimately depend upon the nature of the problem and user preference.

"expressor software is an exciting new vendor in the data integration space. We are very impressed by expressor software's rapid turnaround time in developing a certified data loader for our Netezza appliance."

- Matt Rollender, director of strategic and technology alliances, Netezza

### conclusion

Too often, the desire to implement what should be the correct business solution is easily compromised. Whether this is due to delivery or support costs, skill set availability, timelines, or personal bias – it hinders the ability to choose the optimal architectural solution, resulting in sub-optimal business results and operational stability.

With expressor/Netezza intelligent load & go, the need to compromise on any of your data integration needs are removed – regardless of whether they include data lineage reporting, data quality gates, through to aggregations and data mart generation. Intelligent load & go lets you choose the best data warehouse architecture for your business, thereby helping you to significantly reduce your data warehousing & integration costs.





## biographies



**David Shuttleworth** is a principal consultant at [Edge Associates](#), a partner of Emunio Consulting and expressor software. David has more than 30 years experience in the IT industry, initially in commercial applications programming and systems analysis, then gaining a broader technical background by working for hardware vendors as a systems engineer and consultant. This has given him the opportunity to work with a wide range of programming languages, operating systems, communications products and development tools.

At Edge Associates, David has specialized in the migration of data warehouses to the Netezza platform for major customers such as Debenhams, The Carphone Warehouse, Wandao (now Orange), NTL:Telewest (now Virgin Media) and T-Mobile.

David is a recognized authority in the areas of database technology, parallel processing in the database arena, data warehousing and data mining techniques. He also speaks at conferences and seminars (including creating and delivering the 'Parallel Processing for Commercial Database' seminar for Codd & Date and contributing to articles in 'New Scientist' and BBC Radio 4's 'Network' popular science program). He is a member of the Data Warehousing Institute and has spoken at their annual Leadership Conference and in 2007 gave the Netezza Masterclass presentation at the Netezza European User Forum.



**John Traves** is a principal consultant, non-executive director and co-founder of [Emunio Consulting](#), a systems integrator and reseller partner of expressor software.

John has more than 20 years of international experience in the design and development of information technology solutions, possessing a wide range of software and hardware architecture/disciplines, including more than six years Ab Initio experience. He has provided consultancy to Vodafone, Reuters, O2, a major UK based retailer, Turkcell, Orange and Barclaycard.

After founding Emunio, John built the administration and resourcing functions, developed Emunio's graduate training scheme and managed the implementation of the Emunio HR function. He moved back to the consulting arena, as the company's Principal Consultant in mid 2007 to deliver high level services to clients.



For further information please contact us at:



1 new england executive park  
burlington, ma 01803  
USA  
+1 781.505.4190  
+1 781.505.4191 fax

[www.expressor-software.com](http://www.expressor-software.com)

Copyright 2009 expressor software corporation. expressor, expressor semantic data integration system, smart semantics, intelligent load & go, and redefining data integration are trademarks of expressor software corporation. All other trademarks or trade names are properties of their respective owners. All rights reserved.

PRIVATE AND CONFIDENTIAL