

# expressor<sup>tm</sup>

## redefining data integration

Today's data integration tools are too complex, expensive, inefficient and labor-intensive. They are based on obsolete, CPU-based pricing models and they are difficult to scale. It's time to discover a smarter, faster and lower-cost alternative to the conventional approach to data integration. expressor software tackles the complexity and cost of enterprise IT projects with innovative data integration software that delivers breakthrough development productivity and data processing performance at a significant price/performance advantage.

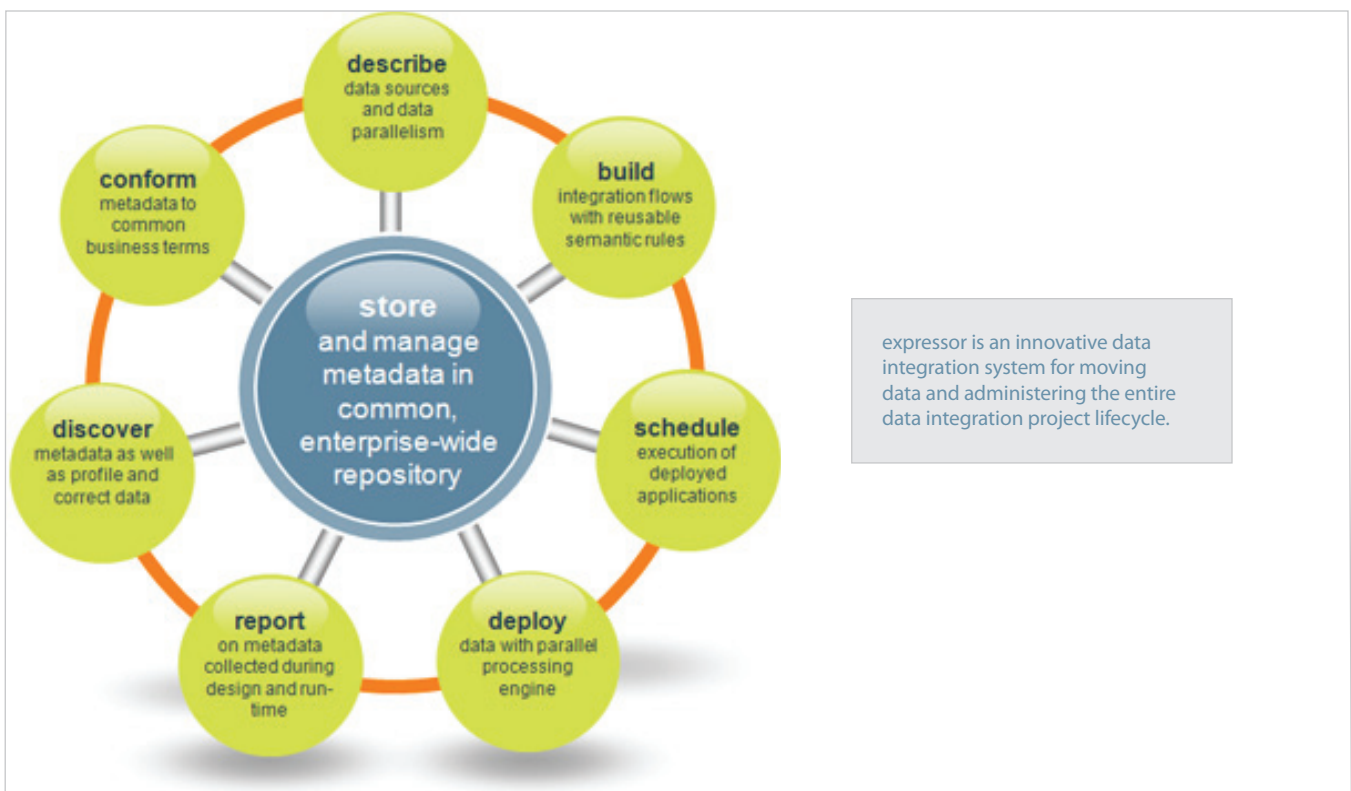
Current technologies force users to hardcode their business rules into the data transformation code, making data integration efforts with these systems very brittle and costly to build and maintain. expressor software is redefining data integration<sup>tm</sup> through a fundamentally new design concept around smart semantics<sup>tm</sup> that addresses crucial pain points in current solutions by simplifying design and reuse via a metadata-driven, repository-enabled system.

## the expressor solution

expressor software's innovative approach allows organizations to rationalize physical metadata constructs around common business terms and write target-specific data transformation and business rules that are 100% reusable across the enterprise, thereby significantly shortening development time and effort. The expressor semantic data integration system™ introduces a new collaborative team development and project management environment built around expressor's semantic metadata repository that offers graphical tools specific to the role and responsibilities of each individual on a project. The environment enables parallel development, facilitates project staging and delivers unparalleled data and process security.

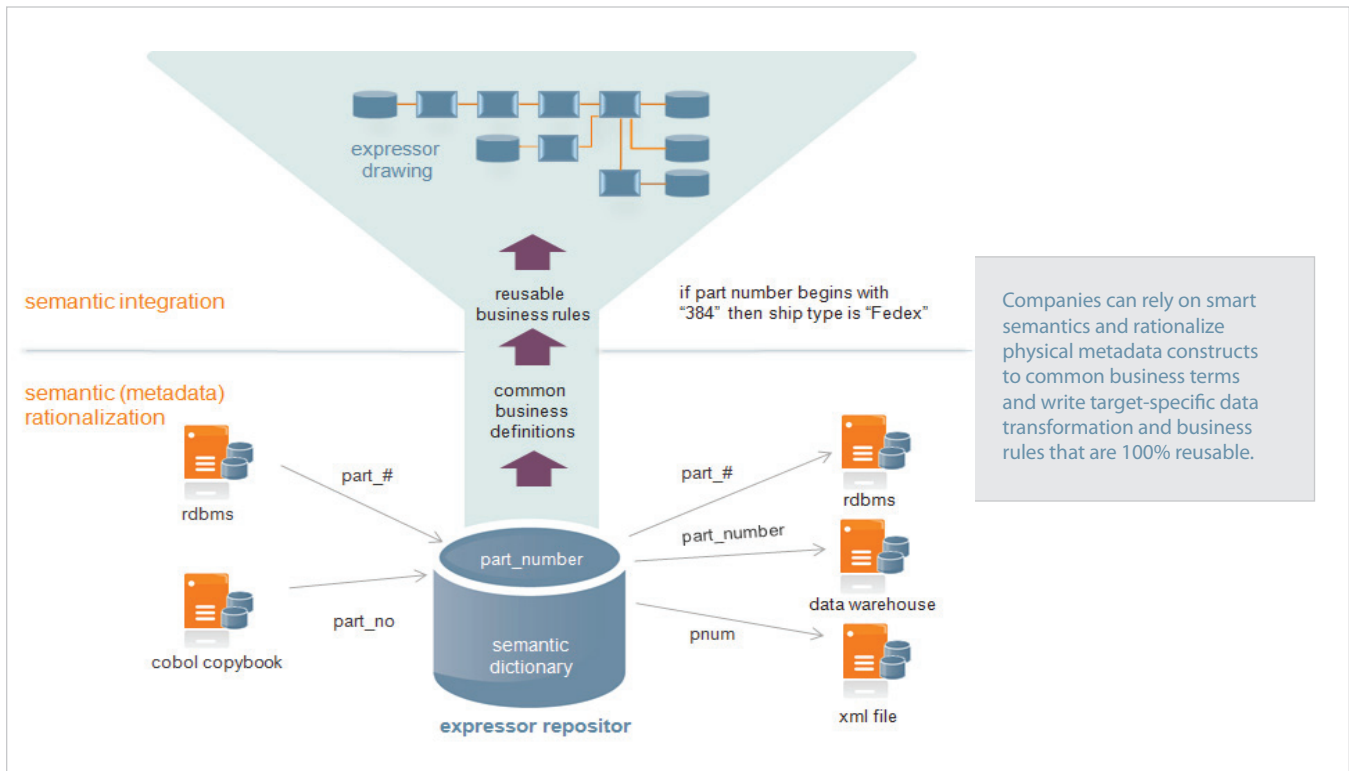
Underlying the expressor system is a high-performance parallel data processing engine that operates in batch or low latency environments and processes terabytes of data per hour, making it the fastest and most scalable engine on the market today. Along with an affordable, usage-based software pricing model, expressor helps companies to greatly reduce their total cost of ownership for data integration. Our ultrafast data processing engine supports parallel execution on heterogeneous platforms without compromise or resource inefficiency. By relying on expressor, organizations benefit from:

- smart semantics: build flexible applications with rationalized business terms and reusable rules
- breakthrough scalability: achieve ultimate data integration performance in batch or real-time
- total lifecycle management: manage your projects with cohesive, role-based integration tools
- reduced total cost of ownership: deliver superior data integration applications for less cost— all the time



### smart semantics

The expressor semantic data integration system addresses crucial pain points in current solutions by simplifying design and reuse via a metadata-driven, repository-enabled system. At the core of expressor is smart semantics, which provides a semantic foundation that enables efficient reuse of business rules and true, collaborative team development so organizations can significantly shorten development time and effort. This allows customers to build flexible projects with rationalized business terms and reusable rules.



Companies can rely on smart semantics and rationalize physical metadata constructs to common business terms and write target-specific data transformation and business rules that are 100% reusable.

Integration design is decoupled from the physical environment, and multiple target-centric business rules can be defined for each business definition. This unique approach to data definition is to semantically correlate data fields only once—from multiple, diverse external information sources—to a set of common names. With smart semantics, organizations can virtualize the external data, providing a consistent, streamlined data integration interface. For example:

- the software automatically recognizes that both the field name “part\_no” in the COBOL copybook source file and the field name “part\_#” in the rdbms target system refer to the same business element.
- expressor maintains affiliations between “part\_no” and “part\_#” by assigning them a common definition (e.g., “part\_number”).
- it automatically accommodates additional field-names references (e.g., “pnum”) as they are encountered in other sources or targets used within the project.
- smart semantics uniquely enables syntactic independence that carries over to subsequent integration projects.

Rules are reusable across projects and expressor ships with built-in business terms for a wide range of data domains across key industries. The smart semantics approach also includes intelligent business definition learning functionality to facilitate the syntactic discovery process for business terms not yet known to the system.

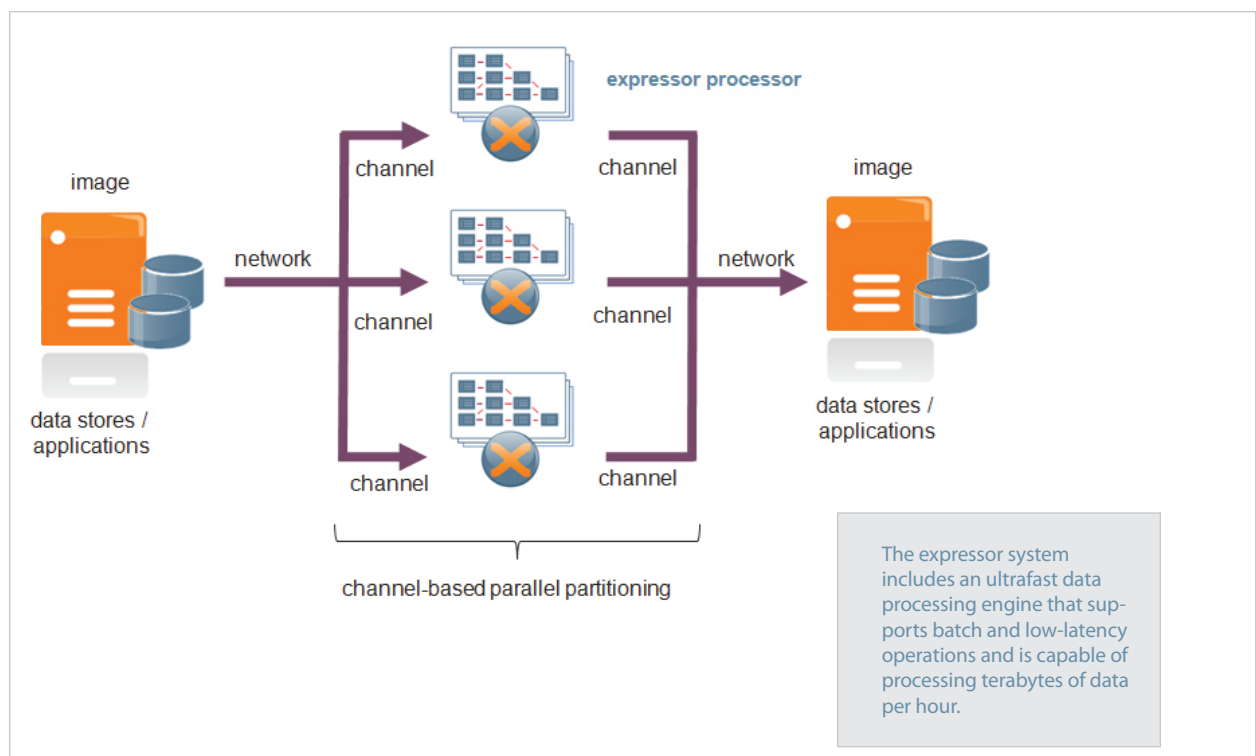
These common definitions are exclusively used within expressor projects, thereby enabling companies to apply consistent business terminology and practices—not only within a single project but also across the organization’s entire data integration landscape. As the foundation for reuse of business rules and many other project artifacts, smart semantics allows customers to improve project efficiency and quality while dramatically accelerating ROI.

## breakthrough scalability

The expressor semantic data integration system provides breakthrough scalability allowing customers to achieve ultimate data integration performance in batch or low latency environments. It is designed and optimized to run on industry leading OS platforms without compromise and delivers resource efficiencies to provide breakthrough scalability.

By abstracting the physical metadata from developers/implementers, expressor provides a simple, common interface for dealing with complex data structures. The expressor system connects to a wide range of external data sources including all major rdbms platforms, SAS applications, messaging queues, ftp servers, flat files, COBOL copybooks and xml files through motors that are configured during the project creation process.

With expressor, a channel is an individual stream of data between tool objects or between an external data source and the application. Channels are used to specify where data is located, to establish parallelism and to identify on which computers the processing should be run; when processing in parallel, there are multiple channels connecting instances of tool processes.



expressor “natively” processes all types and classes of data including complex and hierarchical data and supports a wide range of structured and semi-structured data formats. It is optimized for parallel processing on Windows and Linux platforms. Applications can be moved between environments without modifications, and throughput can be further enhanced by increasing the number of channels executing each operation of the process.

## total lifecycle management

The expressor semantic data integration system enables total project lifecycle management and lets the enterprise manage its projects with cohesive, role-based integration tools. At the core of expressor is an enterprise-wide, centralized repository that offers a single point of reference from which to review, analyze and manage project development and deployment.

The expressor solution poses no restrictions on how development tasks should be performed, whether sequentially or simultaneously. In fact, expressor extends parallelism to the development environment as well as the execution environment. Application developers can begin the design of the integration flow prior to the completion of the data analysis or vice versa. All artifacts from a project—including business rules—are reusable in subsequent projects and stored in the repository, which supports team collaboration and provides version management.

The expressor tools are role-based and targeted for specific tasks typically performed by data architects, data stewards, analysts, implementers and project managers during the integration process. Roles determine privileges, thereby enabling a high degree of data and process security. User activities are carefully tracked, providing a secure and auditable environment.

## reduced total cost of ownership

Central to the expressor solution is the concept of semantic integration, the process of mapping fields from multiple and diverse external data resources to data types and definitions that are then used exclusively within expressor projects.

The input to the semantic rationalization process is the metadata describing the data that will be processed and emitted by the expressor application. The output is an image file and entries within the listings of terms, term abbreviations, and names comprising the dictionary definitions that will be used in subsequent semantic rationalization efforts.

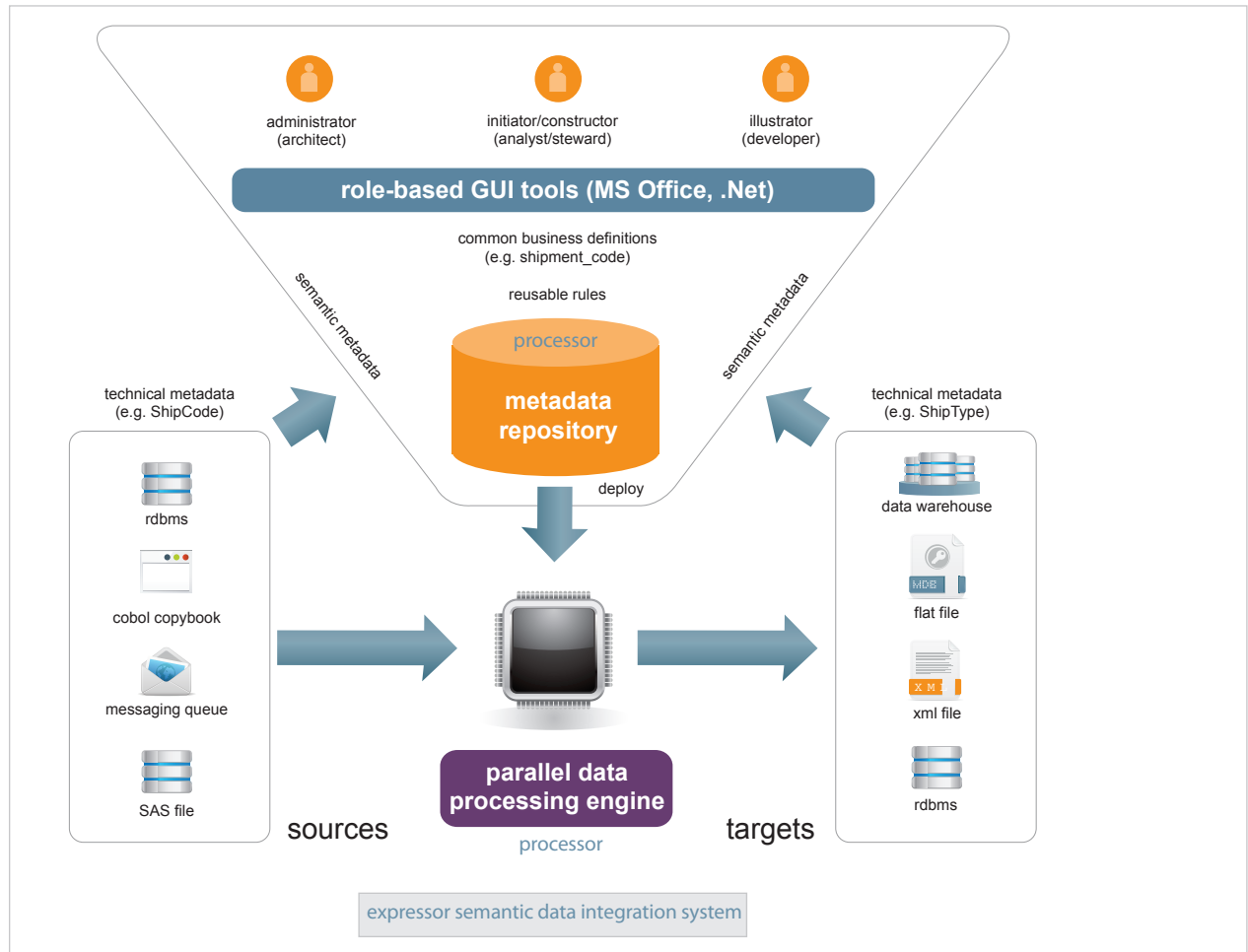
Image files are specific to the resource for which they were written, but different image files will map the same common name — the definition — to differently named fields in their corresponding data resources. For example, the definition “part\_number” may be mapped in two image files to the rdbms table column name “pnum” and to the data item “part\_no” described in a COBOL copybook entry.

The expressor solution represents all data as one of five expressor types — byte, datetime, decimal, number and string. Each definition is associated with an expressor meta type, which is a descriptive alias (e.g., binary or id) for the actual expressor data type (i.e., respectively, byte and string). The meta type assigned to a definition, and its underlying type, may be different from the type assigned to the data in its resource.

The important concept is that individuals developing expressor data integrations work only with the definition names and never need to refer to—or even know—the names used by the actual data resources. Consequently, all of the scripting code and business rule development — referred to as semantic integration — use the definition names, thereby providing a decoupling of the integration design from the physical environment. With expressor, data transformation and business rules are 100% reusable across the enterprise.

## expressor semantic data integration system

With innovative, patent-pending software technology, expressor software tackles the complexity and cost of enterprise IT projects and delivers breakthrough development productivity and data processing performance at a significant price/performance advantage. The company's flagship product — the expressor semantic data integration system™ — has been designed from the ground up based on a semantic metadata foundation.



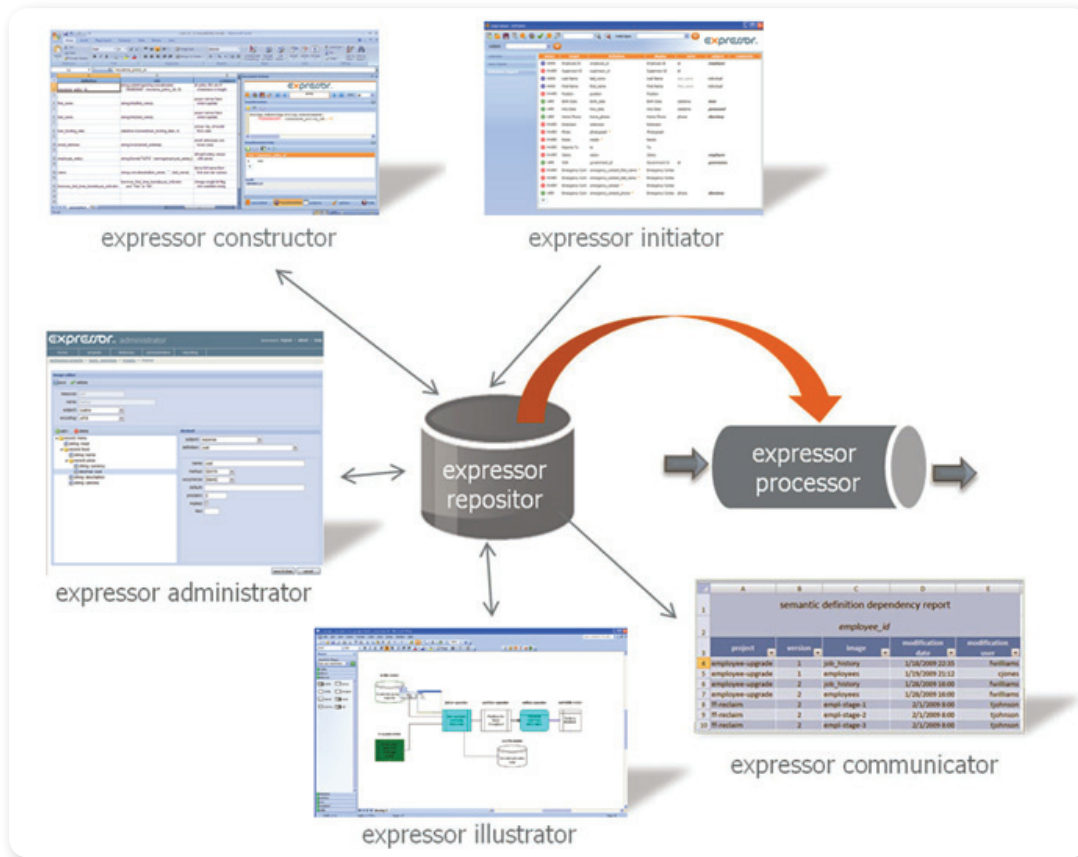
The expressor solution supports Windows and Linux and includes three components that can be installed on a single host or distributed across multiple computers for shared access and optimal performance:

- expressor integrator is a suite of collaborative, role-based tools that support the entire project development and management lifecycle.
- expressor repositor is 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.
- expressor processor is a high-performance parallel data processing engine that runs a deployed data integration application in batch and low-latency environments.

In addition, expressor enables integration with third-party scheduling tools so organizations can leverage OS platform-specific schedulers to invoke expressor processor runtime processes, and expressor investigator provides built-in functions for data quality and error discovery, and offers interfaces to third-party data quality tools.

## expressor integrator

With expressor integrator, organizations benefit from collaborative, role-based tools that support the entire project development and management lifecycle. The tools include web interfaces for project setup, administration, and role-based reporting, and Windows desktop applications for defining reusable rules and integration flows. The expressor integrator applications support total lifecycle management, and they are seamlessly integrated with the expressor repositor database and version control system.



The expressor integrator provides built-in project and data security according to a user's role on a project, and it includes the following:

### expressor illustrator

expressor illustrator is a Windows desktop-based visual integration flow design application geared at ETL developers who are responsible for designing and testing a data integration application. It relies on image and network files created within expressor administrator, and is implemented as a plug-in for Microsoft Visio that provides a drag-and-drop construction interface using pre-built shapes.

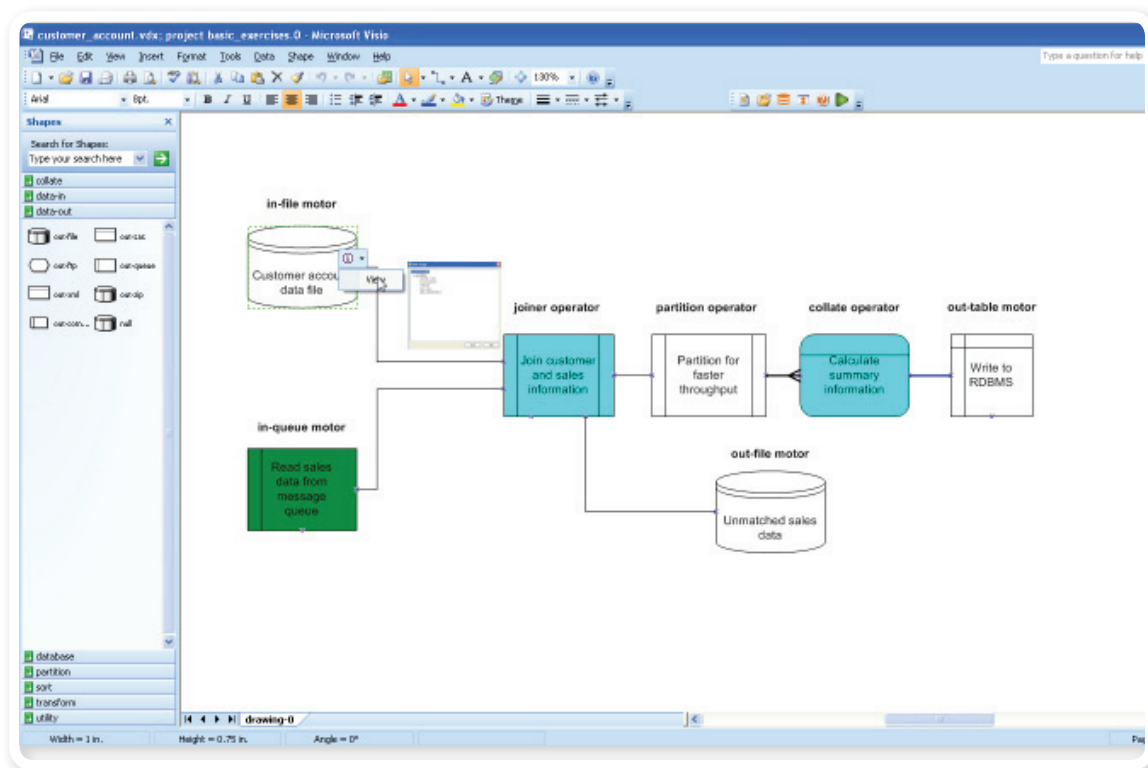
A developer begins by checking out a project and downloading the relevant image and network files that describe data sources, record structures, and parallel processing methods. They then visually create data integration projects consisting of one-or-more expressor drawings, using the pre-built shapes and by writing data processing instructions using expressor datascript™, a powerful scripting language based on the widely adopted Lua language with expressor extensions.

expressor drawings are assembled from a collection of prewritten components that perform tasks such as:

- partitioning, filtering, sorting, and collating records
- reading from—and writing to—relational databases, flat files, xml files, ftp servers, SAS files, or messaging queues
- transforming the format and content of data records

expressor illustrator also features a powerful point-and-click interface for the collate operator that allows standard summary calculations to be easily configured in only a few steps. Users can automatically generate code for summary calculations—such as for calculating the standard deviation within a group of records; more complex calculations can be developed by writing procedural processing logic in datascript.

expressor illustrator supports remote development, allowing a user to run a drawing under development on a remote expressor processor server for testing purposes.



expressor illustrator allows developers to leverage a flexible user interface to drag-and-drop pre-built shapes to build an integration flow.

Developers can write business and transformation rules from within expressor illustrator that define the processes to be applied to the data when creating the fields in an output record. Developers can also test their processing logic using synthetic debugging, a mechanism to exercise the rules in expressor illustrator with test values, without connecting to the physical data.

expressor illustrator provides feedback during the development process through the use of color-coding and a messages window. Complex data transformations can be described through the expressor datascript language, and scripts can be either embedded directly in the drawing shape or written in an external file and called from embedded scripting statements. A developer can select from a pick-list of existing functions or create new ones, which can be shared across the enterprise.

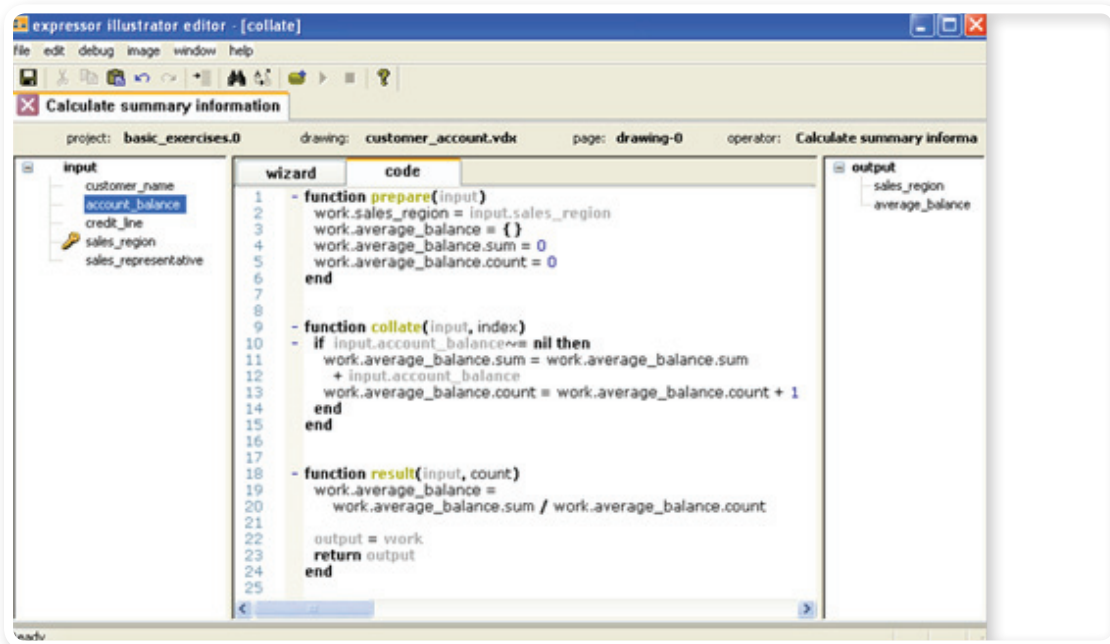


Traditional data integration solutions on the market are complex, bloated and service-intensive. Organizations can now tackle the complexity and cost of enterprise IT projects with semantic data integration software that delivers breakthrough development productivity and data processing performance at a significant price/performance advantage.

The expressor solution allows companies to reuse the most labor-intensive elements of data integration projects, deliver superior data processing performance and reap significant cost savings in development project lifecycle management and in hardware and software licensing. The following are just a few of the solutions enhanced by the expressor system:

- affordable data migration
- high-performance data warehousing
- rapid ETL migration to expressor
- data governance
- low-latency data processing
- complex XML data processing
- SSIS performance alternative
- complex data transformations
- intelligent load & go
- operational data integration

While developers do not need to view the actual data, the network file may still point to production or historical test data, which can then be used in developing and debugging the application. When the application runs, the developer cannot view the data unless the developer is assigned a privileged role that allows for data viewing. Roles-based user segmentation provides organizations with maximum flexibility for controlling access and security to critical enterprise information.



expressor illustrator includes an editor where developers can leverage the datascript language to define the business logic.

## expressor administrator

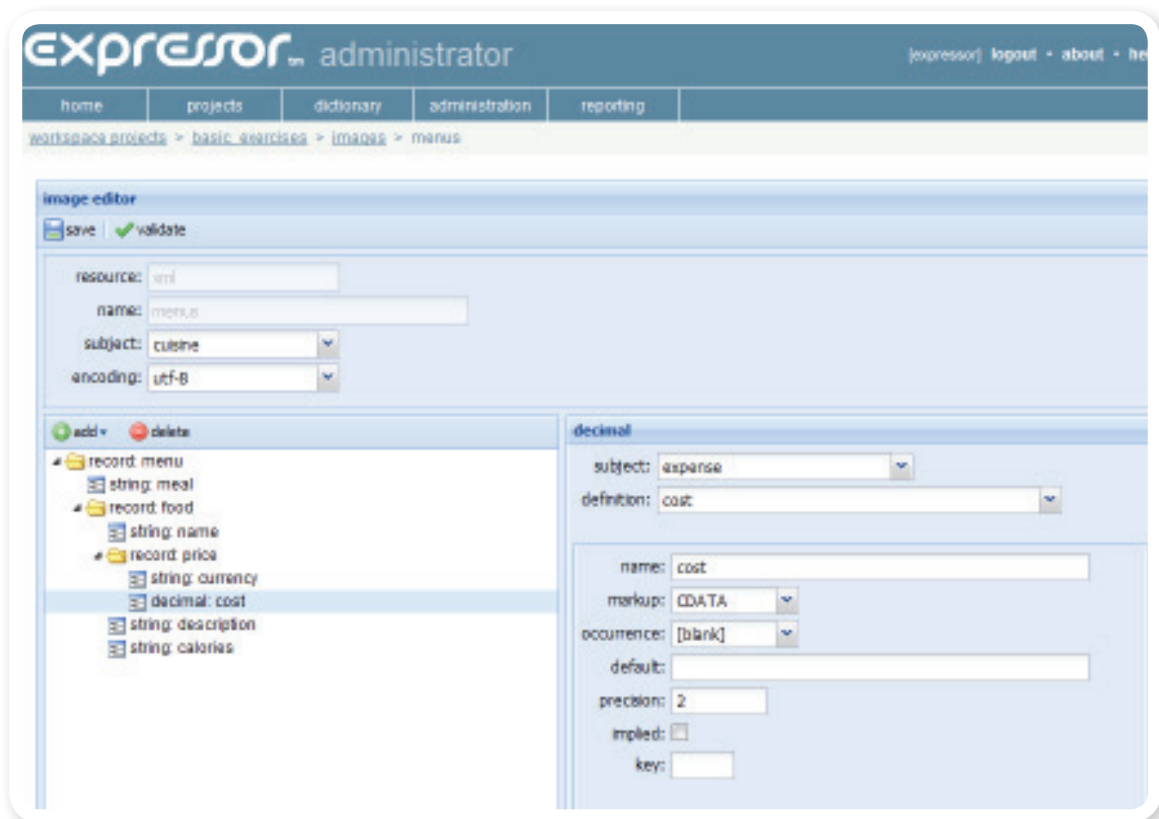
expressor administrator is a comprehensive web-based project setup and management application. Architects, data stewards, administrators, and project managers use expressor administrator to register hardware and software resources with which an expressor drawing will interact, set up an expressor project, manage personnel assignments, perform semantic rationalization tasks, and create the data and network descriptions used by developers. It is the primary interface for setting up a project and creating:

- image files—xml files that provide descriptions of external data record structures
- network files—xml files that identify where data or processing is located and the level of parallelism deployed
- role assignments—user access privileges within the expressor environment

Image and network files provide the foundation for projects, and role assignments determine user privileges. Role-based privileges allow organizations to streamline productivity and simplify integration. For example, the role of architect has the most privileges (but not all) within expressor, the role of developer is limited to use of expressor illustrator, and the role of manager only has privileges for assigning users to projects.

An architect would create a project in expressor administrator and write the required image and network files. Then the architect or manager would assign the users participating in the project.

In expressor, whenever you define a term, business definition, or image file, it is associated with a subject area, which is a category for grouping artifacts. expressor administrator ships with pre-defined subjects including banking, account, and country, and provides the ability to manage subject areas. Users can easily add new subjects and de-activate pre-defined subjects that do not apply to their data integration project.



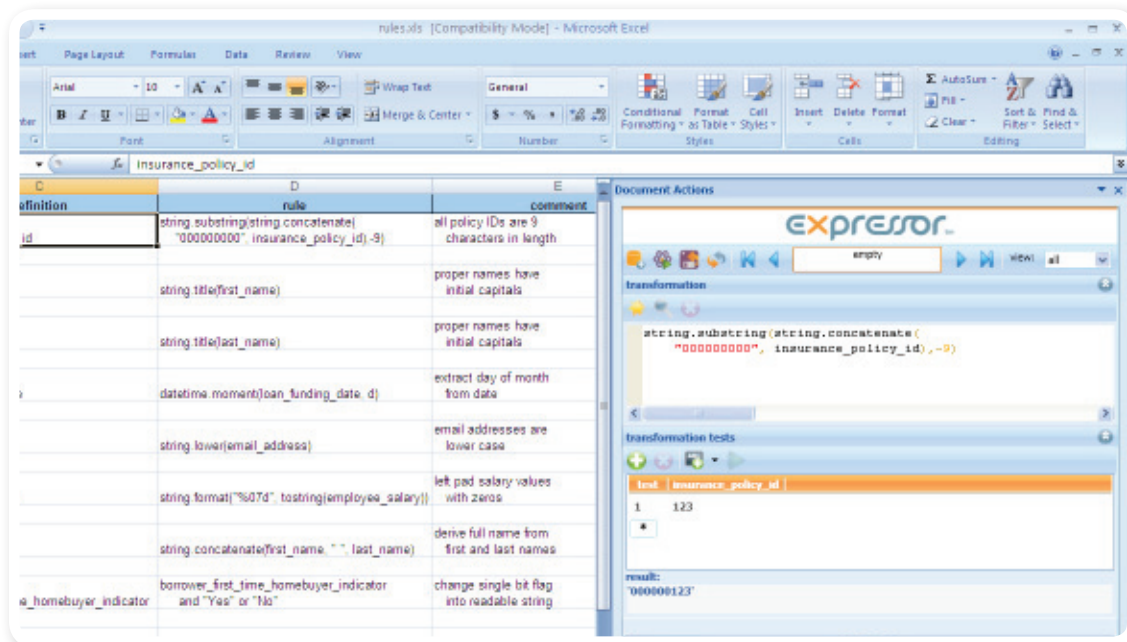
expressor administrator provides a friendly graphical user interface that allows users to easily create images of data structures.

expressor administrator features a user-specific workspace for building project image and network files. This new model follows the same check-out/check-in paradigm as used in expressor illustrator and expressor constructor, and it allows for networks and images to be developed over time and over many user sessions before being committed to the expressor repositor.

expressor administrator offers the ability to copy images, networks, and projects as well as the ability to create project version snapshots that preserve published project state and can then be used through the expressor communicator for historical analysis. It also provides the ability to automatically create an image from relational tables and COBOL copybooks, as well as a complete API to allow import of any other metadata descriptions.

## expressor constructor

expressor constructor is a Windows spreadsheet-based desktop application targeted at data stewards and data analysts tasked to define semantic definitions and describe simple business rules based on these definitions. It has an interface integrated with Microsoft Excel that allows users to easily add metadata and business rules. For convenience, expressor constructor has both a spreadsheet interface and a wizard interface.

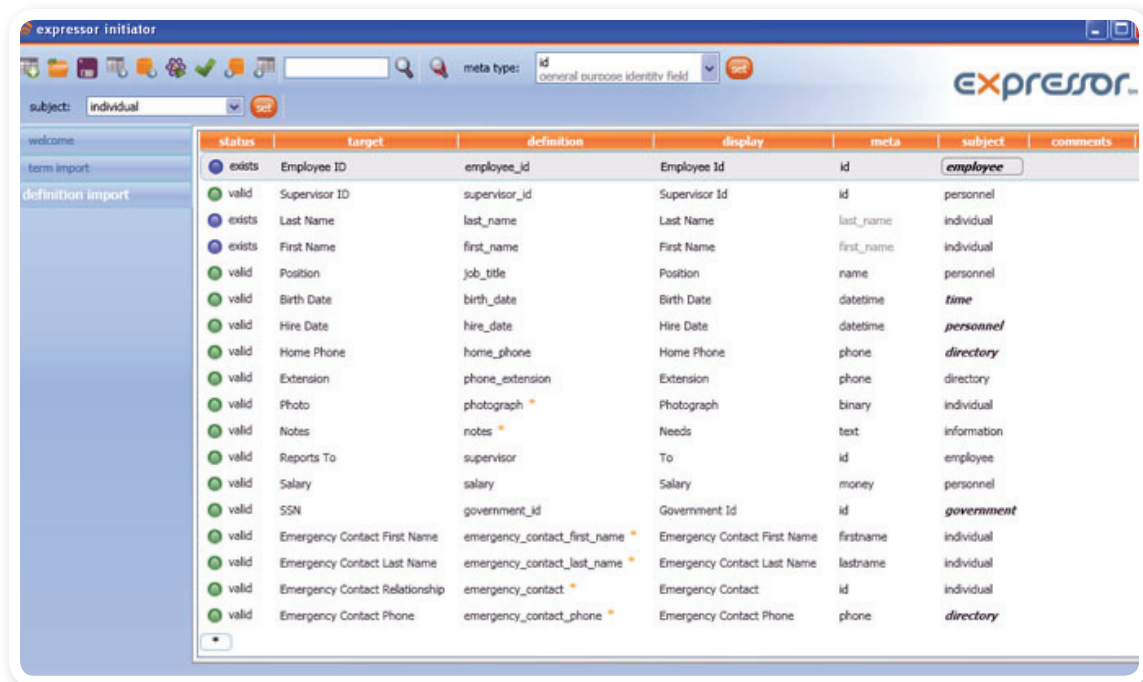


With expressor constructor, you can easily describe business rules that are portable between integration efforts.

For example, a user can add or import field names and the wizard will suggest rationalized names. Once the rationalized names are selected, the user can write business rules. This architecture provides a virtualization layer so data analysts can concentrate on creating and validating the business rules without being distracted by the complexities of the external data structures. It allows data analysts to define target-centric business rules that can be checked-into expressor repositor and shared across projects. These rules are then automatically suggested to the ETL developer in expressor illustrator.

## expressor initiator

expressor initiator is a Windows desktop application that allows data stewards to rapidly populate the semantic dictionary in expressor repositor. expressor initiator is especially useful when a large number of terms and business definitions must be added to the dictionary because it can process a collection of entries from external spreadsheets or database tables in a single batch.



expressor initiator allows data stewards to import terms and definitions from external sources in bulk and rationalize them against existing definitions in the semantic dictionary.

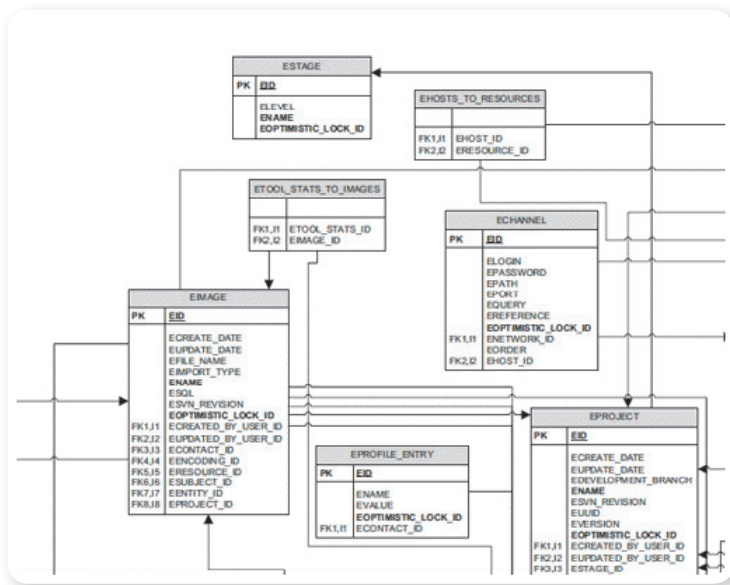
expressor initiator includes built-in syntax checking and data validation, and it will rationalize newly imported business definitions against definitions already in the dictionary. When a new definition is encountered, data stewards can complete the description through an easy-to-use grid interface rather than extensive text entry. Because this new productivity tool allows semantic definitions and terms to be created and edited in bulk, users can complete the description of multiple definitions in a single operation.

Since input to expressor initiator can be an Excel spreadsheet or a connection to a database table, individuals unfamiliar with the complete suite of expressor integrator tools can still contribute to the process of initializing the expressor dictionary and building support for the semantic rationalization process.

## expressor communicator

expressor communicator is a reporting database that provides visibility into application projects and operations. It works with JDBC-based or ODBC-based off-the-shelf reporting tools and interfaces, and allows users to build custom reports for impact analysis, dependency tracking, dictionary and artifact usage, and operational statistics. For example, users can even leverage Microsoft Excel to swiftly create reports by running queries against expressor communicator.

expressor communicator consists of a read-only database optimized for report writing that mirrors the information in the expressor repositor database, allowing organizations to benefit from a separate reporting database. It exposes information stored in expressor repositor so that administrators can easily create queries to determine project status without degrading performance.



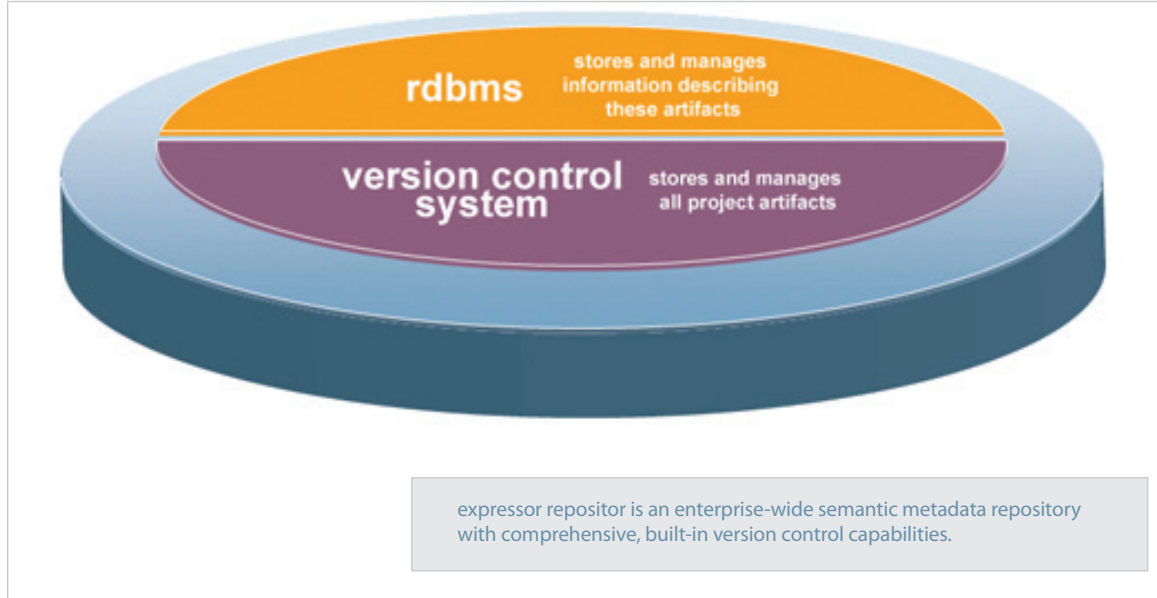
The expressor communicator reporting database provides visibility into application projects and operations.

expressor communicator also provides a number of sample SQL report queries that cover a wide range of common reporting tasks and act as templates for developing custom reports. It currently includes 20 pre-written queries that simplify and automate reporting, such as “who created one of these items”, “who changed one of the items”, “report when a drawing was run”, and “list all the images defined in a project.”

### expressor repositior

The expressor repositior is an enterprise-class semantic metadata repository that collects, stores, and manages project information, reusable data descriptions, application file versioning, performance metrics, and the implementation and enforcement of role-based security. It enables total lifecycle management and centrally maintains the details of a data integration application, including user roles and assignments.

It maintains descriptions of the incoming and outgoing data record structures, parallel processing paradigms, and business rules, providing the information needed to manage project development and support smooth analysis and management of data processing tasks.



expressor repositior is an enterprise-wide semantic metadata repository with comprehensive, built-in version control capabilities.

expressor repositior has a powerful new paradigm for version control that allows users to work on the same artifact, and it will warn users of unknown changes and force resolution. It allows users to check-in and check-out resources while providing a centralized audit trail. expressor repositior now supports enhanced rule capture capabilities, incorporating business rules created in expressor illustrator by developers.

Business definitions, business rules, and host computer descriptions can be reused in multiple projects, reducing the development effort and ensuring adherence to corporate data standards and business rules. expressor repositior offers a flexible configuration and parameterization architecture that supports:

- dynamic execution
- migration between environments
- construction of reusable components

## expressor processor

The expressor processor is a high-performance parallel data processing engine that runs a deployed data integration application. The expressor processor can be deployed in a mixed environment composed of computers running:

- Windows
- Linux

The expressor processor can run in either intermittent or perpetual mode, supporting both batch and low-latency, real-time processing. It provides extensive connectivity to a wide range of data sources, including:

- ftp servers
- relational databases
- xml files
- complex flat files including hierarchical and multi-format files
- SAS files
- messaging queues

expressor has been certified against all major relational database systems including Oracle, IBM DB2, Microsoft SQL Server, and Sybase and ships with database drivers for a variety of other database management systems including Informix, PostgreSQL and MySQL. It also supports data warehousing and business analytics platforms (data warehouse appliances) from Netezza and Teradata, and includes ODBC support to enable integration with other data sources and targets.

expressor provides comprehensive support for complex data formats included in flat files and xml files. Unlike other products, the expressor processor preserves the rich complex physical structures in the data during processing. expressor provides support for messaging queue motors to integrate with MQ Series and Microsoft MQ messaging systems. expressor processor is available on Windows and Linux RedHat platforms.

expressor processor includes many key parallel data processing features, performance optimizations, and exception handling capabilities. For RDMBS exceptions that occur during out-table operations (such as constraint violations), unprocessed out-table records can be redirected to an alternate output port and captured for custom post-processing and recovery. Similarly, with messaging systems, incoming messages with mis-formed content will not be submitted for processing, but redirected to an output port and captured for further analysis and review.

## native parallelism key to achieving high throughput and performance

One of the primary key features of expressor is its patent-pending parallel-based systems architecture and runtime environment. Unlike many of the currently “mainstream” ETL and data integration solutions on the market that were retrofitted with acquired parallel technology or had parallel processing capabilities introduced as an afterthought into their core architecture, expressor software started from scratch to build the highest performing and most highly scalable parallel-based data integration solution on the market.

expressor processor includes the following key features that enable the highest throughput and performance:

- native built-in parallelism
- implicit and explicit parallelism
- depth parallelism
- data pipelining
- data partitioning
- channel based parallel partitioning

expressor natively supports all these features of parallelism in a design and runtime environment that is easy to learn, use, manage, and monitor. expressor processor can process terabytes of data per hour and favorably competes against all leading parallel data processing systems on both performance and price.

## find out how expressor is redefining data integration

expressor software provides next-generation data integration software based on a unique semantic metadata foundation. The company was established in 2007 by experienced data integration and data warehousing practitioners and executives. The company provides smart, ultrafast, and affordable data integration software, is headquartered in Burlington, MA and funded by Commonwealth Capital Ventures, Globespan Capital Partners and Sigma Partners.

Version 2.0 of the expressor semantic data integration system shipped in July 2009 and expressor 2.2 will be released in April 2010. expressor 2.2 includes an option for automatic semantic rationalization, a new enhanced connectivity framework, team development improvements, and significant enhancements in support of Microsoft’s Windows, Office and SQL Server platforms.

The enhancements in version 2.2 of expressor were based on extensive customer and partner feedback during rigorous proof-of-concept benchmarks and project deployments, as well as new requirements stemming from a wide variety of analytical and operational data integration use cases.

expressor targets high-performance data integration opportunities in mid-size SMBs and departmental organizations within global 2000 enterprises. The company is developing an extensive partner network consisting of technology, reseller and system integration partners in the US and EMEA.

expressor software corporation  
1 new england executive park  
burlington, ma 01803  
usa

+1 (781) 505-4190 tel  
+1 (781) 505-4197 fax

© 2010 expressor software corporation, all rights reserved. The following are trademarks of expressor software corporation: expressor, expressor semantic data integration system, smart semantics, intelligent load and go, expressor datascript and redefining data integration. All other trademarks are properties of their respective owners.

exp-po-0410

**EXPRESSOR**™  
redefining data integration

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