SAS® Enterprise Data Integration Server

A complete solution designed to meet the full spectrum of enterprise data integration needs

What does SAS® Enterprise Data Integration Server do?

SAS Enterprise Data Integration Server, featuring DataFlux® technology, is a powerful, configurable and comprehensive solution that can meet a wide variety of data integration requirements, from small tactical projects to strategic business initiatives. It can:

- Access virtually all data sources.
- Extract, cleanse, transform, conform, aggregate, load and manage data.
- Support data warehousing, migration, synchronization, federation and provisioning initiatives.
- Support both batch-oriented and real-time master data management solutions.
- Create real-time, reusable data integration services in support of service-oriented architectures and data governance.

Why is SAS® Enterprise Data Integration Server important?

It enables organizations to efficiently manage data integration projects on an enterprise scale in a timely, cost-effective manner and meet the high data quality expectations of information consumers.

For whom is SAS® Enterprise Data Integration Server designed?

It is designed for organizations in all industry sectors that are implementing one or more data integration projects, dealing with changing business landscapes and business-driven IT initiatives, trying to meet regulatory requirements, or implementing data governance.

Organizations struggle daily with the challenges of distributed and rapidly increasing data volumes, inconsistently defined data across disparate IT systems and the high expectations of data consumers across the enterprise who depend on information to be correct, complete and available when they need it.

SAS Enterprise Data Integration Server provides a comprehensive solution that enables organizations to solve these challenges in a timely, cost-effective manner with the ability to efficiently manage data integration projects on an enterprise scale – increasing overall productivity and reducing the total cost of ownership.

Key Benefits

- **Always access the data you need.** From older legacy systems to the latest ERP applications, data from virtually any hardware platform or operating system can be accessed, cleansed and processed. New source systems can easily be added and security is managed centrally. This saves time, shortens learning curves and gives decision makers the complete information they need.
- **Leverage/reuse work by others.** A common repository enables the centralized storage, management and reuse of work based on user authorizations, reducing both development and maintenance time.
- **Improve productivity.** A GUI environment that is easy to use provides a standard interface for building and documenting work. Collaboration is encouraged and manual coding is available when needed. New team members can get up to speed quickly on work done by others, which is important when documentation is inadequate or missing.
- **Manage security and administration at all levels.** Reusable templates make it quick and easy to provide role-based authorizations and administrative privileges at the user, department or enterprise level.
- **Meet time constraints even within decreasing windows of availability.** SAS processes data fast! Organizations can take advantage of a grid-enabled load-balanced, multithreaded parallel processing architecture that can quickly transform and move data between different platforms and systems.
- **Deliver consistent, trusted and verifiable information.** Consistently getting correct data when and where it is needed provides increased confidence in the accuracy and timeliness of operational and business information. Data quality auditing tools monitor the quality of data in processes and source systems. Users can see where data originated and how it was transformed. Optional enrichment components can add value and ensure everyone receives the best possible data.
- **Eliminate overlapping, redundant tools and systems with one solution.** SAS offers the only comprehensive enterprise data integration solution that is built from the ground up to meet the full spectrum of data integration needs. It eliminates the piecemeal approach of linking and managing technologies from different vendors and provides lower overall cost, reduced risk and faster results.

Product Overview

SAS offers the only comprehensive enterprise data integration environment that is built from the ground up to meet the full spectrum of enterprise data integration needs. Instead of linking and managing technologies from different vendors, SAS Enterprise Data Integration Server provides a collaborative design environment promoting object reuse and sharing, administrative controls, wizard-driven design process workflow, and ease of use and maintenance. This flexible, reliable solution can access data from virtually any system in any form, transform and cleanse data even in real time, and handle data migration, synchronization and federation projects all through a versatile services environment that is easy to deploy and maintain.

Interactive Data Integration Development Environment

A graphical user interface (GUI) simplifies and speeds projects with wizards, extensive built-in transformations and powerful productivity enhancements, all while providing a single point of control for managing complex enterprise data integration processes. SAS Data Integration Studio is easy to learn and use. It provides a collaborative environment that lets you build reusable processes to speed data integration development both now and in the future. It automatically captures and manages standardized metadata from any source, and enables you to easily display, visualize and understand enterprise metadata and your data integration processes.

Connectivity and Data Access

Most organizations struggle with accessing the plethora of data sources (legacy, relational, flat files, XML, cloud data, text, etc.) that are necessary to support analytical systems. SAS Enterprise Data Integration Server provides connectivity to virtually all types of data sources and types, operating systems and hardware environments using both native access and open standards. It also supports the reading and writing of data from message queues and the sending and receiving of data to and from Web services.

Metadata Management

SAS provides a shared metadata environment that is both independent (for data integration) and part of SAS' comprehensive platform. Technical, business, process and administrative metadata is stored and managed in a way that leverages and facilitates reuse of existing table definitions, business rules and more. Navigational tools help users understand how the data was derived and where it is stored and used. Shared metadata provides a consistent definition across data sources to speed integration projects, simplify design and reduce maintenance costs.

Data Cleansing and Enrichment

There is an increased awareness, driven by compliance mandates and data breaches, of how data quality and data governance can directly affect the bottom line. This puts increasing pressure on IT to address potential data quality issues. SAS Enterprise Data Integration Server provides a single environment that seamlessly integrates data quality within the data integration process, taking users from profiling and rules creation through execution and monitoring of results. From data deduplication (for example, within database marketing applications) to cleaning up data (for example, before storing in a data warehouse), SAS provides an enterprise approach that lets you develop and share a library of data rules and processes between projects and across the entire data integration solution. Organizations can transform and combine disparate data, remove inaccuracies, standardize on common values, parse values and cleanse dirty data to create consistent, reliable information.

Extraction, Transformation and Load (ETL) and Extraction, Load and Transformation (ELT)

Loading data warehouses and data marts within their allotted time windows, quickly building analytical marts for special projects, and creating extract files for reporting and analysis applications are tasks IT organizations face each day. SAS Enterprise Data Integration Server includes an intuitive point-and-click Design Editor window that allows developers to easily build logical process workflows, quickly identify the input and output data stores and create business rules in metadata, enabling the rapid generation of data warehouses, data marts and data streams. Users can also choose to have many transformations and processes take place inside a connected database, data warehouse or storage system. This is referred to as ELT, push-down or in-database processing, and can substantially speed up overall processing times by reducing unnecessary data movement. SAS Enterprise Data Server uses visual SQL push-down to select the optimal processing approach.

Migration and Synchronization

Moving data from system to system is a constant activity in most organizations. Mergers and acquisitions result in multiple, overlapping systems containing information that often needs to be synchronized and ultimately

migrated. Moving legacy data during upgrades and conversions is an ongoing process, as is the movement of data into and out of ERP systems.

SAS Enterprise Data Integration Server provides the capability to migrate, synchronize and replicate data across different operational systems and data sources. The point-and-click process design editor makes it easy to document migration and synchronization processes in workflows that can be reused and modified for other projects. Powerful data transformations are available for altering, reformatting and consolidating information during these processes.

You also can build a library of reusable business rules ensuring that bad data is never spread from system to system. In this way, information delivered across all applications, systems, environments and geographies is up-to-date, consistent and accurate.

Data Federation

Organizations today have data stored and scattered in and across numerous data sources. Often what's needed is fast access to the most current operational data to support real-time analytics and reporting needs. SAS Enterprise Data Integration Server provides the ability to query and use data across multiple systems without the physical reconciliation or movement of source data. The logical semantic layer shields business users from the complexities of the underlying physical data. By avoiding unnecessary data replication and movement, it is possible to quickly and cost-effectively deliver up-to-date data that is consistent and accurate.

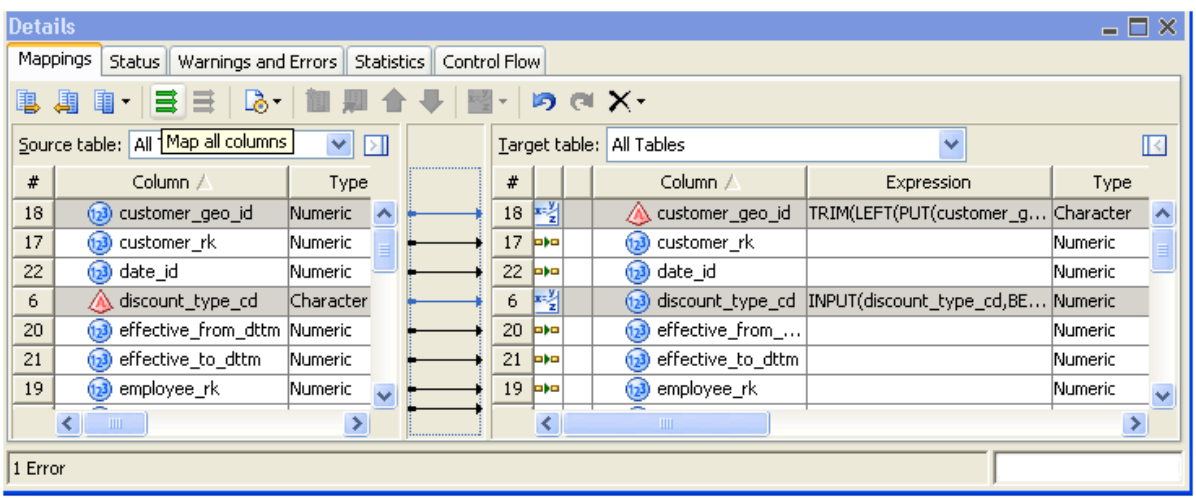
Master Data Management Support

SAS Enterprise Data Integration Server includes data-mastering capabilities that provide a basis for implementing master data management projects that enable you to identify, standardize and correct common master data such as customer and product data. Unsurpassed data access, profiling, enrichment, clustering and consolidation to clean, standardize and enhance data, and an intuitive development environment that is adaptable to each organization's technologies and standards, increase productivity and produce more rapid results.

SOA and Message Queue Integration

Organizations are challenged to improve operational efficiency, streamline processes and be more agile. Using a service-oriented architecture (SOA) approach helps IT ensure that various applications can communicate with each other to better meet changing business requirements. SAS Enterprise Data Integration Server delivers easily maintained data services that enable developers to build sophisticated data services once and deploy them across the enterprise for reuse.

Message queue integration is another way to reduce maintenance, integration costs and bridge new technologies. Using SAS Enterprise Data Integration Server, you can access message queues in batch or real time without the need for custom programming. Integration developers simply treat message queues as any other source and target.

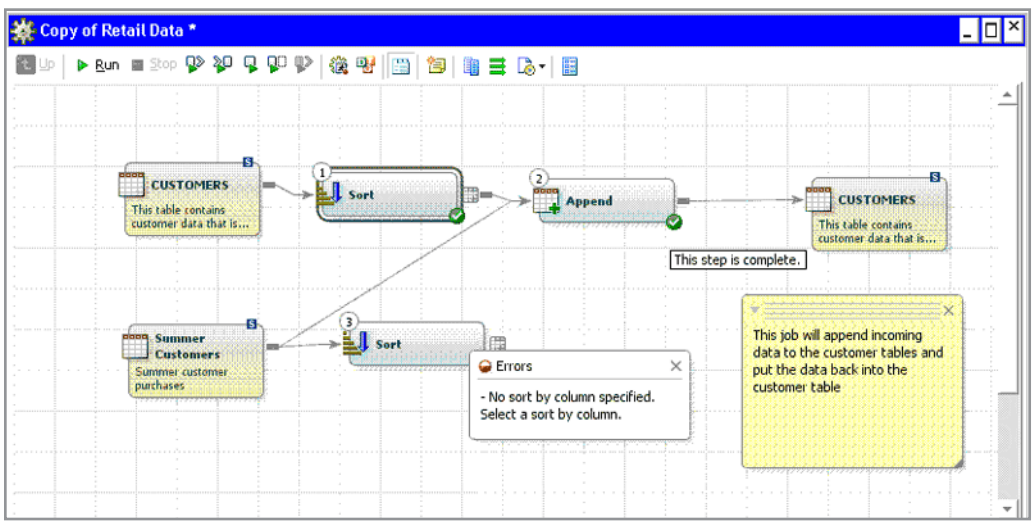


With SAS Enterprise Data Integration Server, you can define the propagation of information from table to table in your transformations. This shows an example of the default mapping rules being applied when mapping numeric to character columns, and character to numeric columns. This also shows the intelligence of the mapping display that uses color to indicate the presence of a transformational expression between source and target. Default mapping rules are predefined but can be extended to meet business-specific needs.

Key Features

Interactive data integration development environment

- An easy-to-use, point-and-click GUI uses an intuitive set of configurable windows for managing data integration development processes.
- A visual, end-to-end process designer lets developers quickly build and edit processes.
- Drag-and-drop functionality eliminates programming.
- Wizards make it easy to access source systems, create target structures, import and export metadata, and build and execute ETL process flows.
- The multiple-user, multiple-level design environment supports collaboration on large, enterprise projects.
- Role-based permissions show users only what they are authorized to see.
- Customizable metadata tree views let users display, visualize and understand metadata.
- Dedicated GUI for profiling data to identify and repair source system issues while retaining the business rules for use in other ETL processes.
- Interactive debugging and testing of jobs during development and full access to logs.
- Check-in/check-out of jobs, related tables and objects; and job status viewing.
- Audit history lets designers see which jobs or tables were changed, when and by whom.
- Ability to distribute data integration tasks to nearly any platform and to connect virtually any source or target data store.
- Integration with third-party vendors Subversion and CVS provides enhanced version and source control features such as archiving, differencing and rollback.
- Job status and performance reports provide the ability to track metrics such as CPU use, memory, I/O, etc.
- Automated job deployment allows the use of common scripting languages to deploy SAS Data Integration Studio batch jobs in an automated manner.
- Enhanced SAS code import capabilities give current SAS users an easy way to import their SAS jobs and SAS code into SAS Data Integration Studio. Includes improved logging and error checking.
- New command-line job deployment for deploying single and multiple jobs.
- Enhanced data integration job orchestration (process flow).
- The ability to surface in-database scoring models within SAS Data Integration Studio.
- Enhanced connectivity to Aster Data, EMC Greenplum and Sybase IQ databases with the ability to push down more processing to the databases.
- Native support from within SAS Data Integration Studio to Oracle's bulk-loading capabilities.



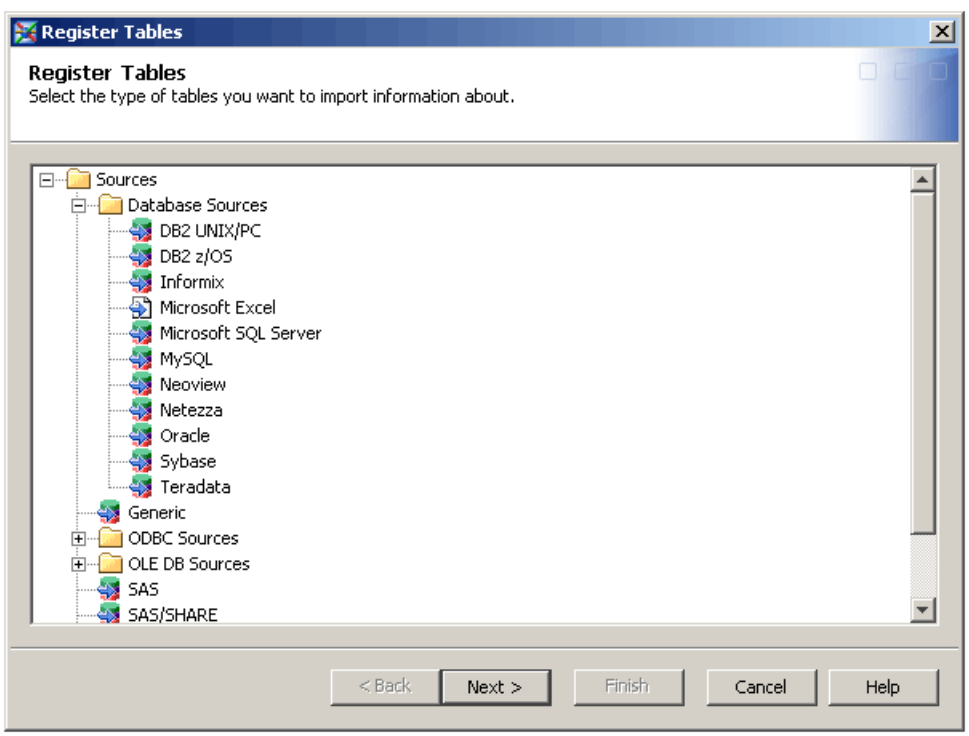
SAS Enterprise Data Integration Server includes an easy-to-use and informative GUI. You build jobs by dragging and dropping data objects into the diagram area. You can add transformations such as sorts, joins and loads from a library and draw arrows to connect the objects together. Self documentation is provided using annotated data, and yellow notes containing further information can be added by users.

Connectivity and data access

- Provides connectivity in batch or through message queues in real time to more data sources on more platforms than any other solution.
- Data access engines are available for enterprise applications, nonrelational databases, RDBMSs, data warehouse appliances, PC file formats and more.
- A complete and shared metadata environment provides consistent data definition across all data sources.
- Native access methods deliver the best performance and reduce the need for custom coding.
- Support for message-oriented middleware, including WebSphere MQ from IBM, MSMQ from Microsoft, Java Message Service (JMS) and Tibco's Rendezvous.
- Support for unstructured and semi-structured data to parse and process files.
- Access to static and streaming data for sending and receiving via Web services.
- Expanded support for MPP databases: AsterData nCluster, EMC Greenplum and Sybase IQ, enabling more ELT pushdown and support for bulk-load utilities.
- Native support for SQL-based processing.

Metadata management

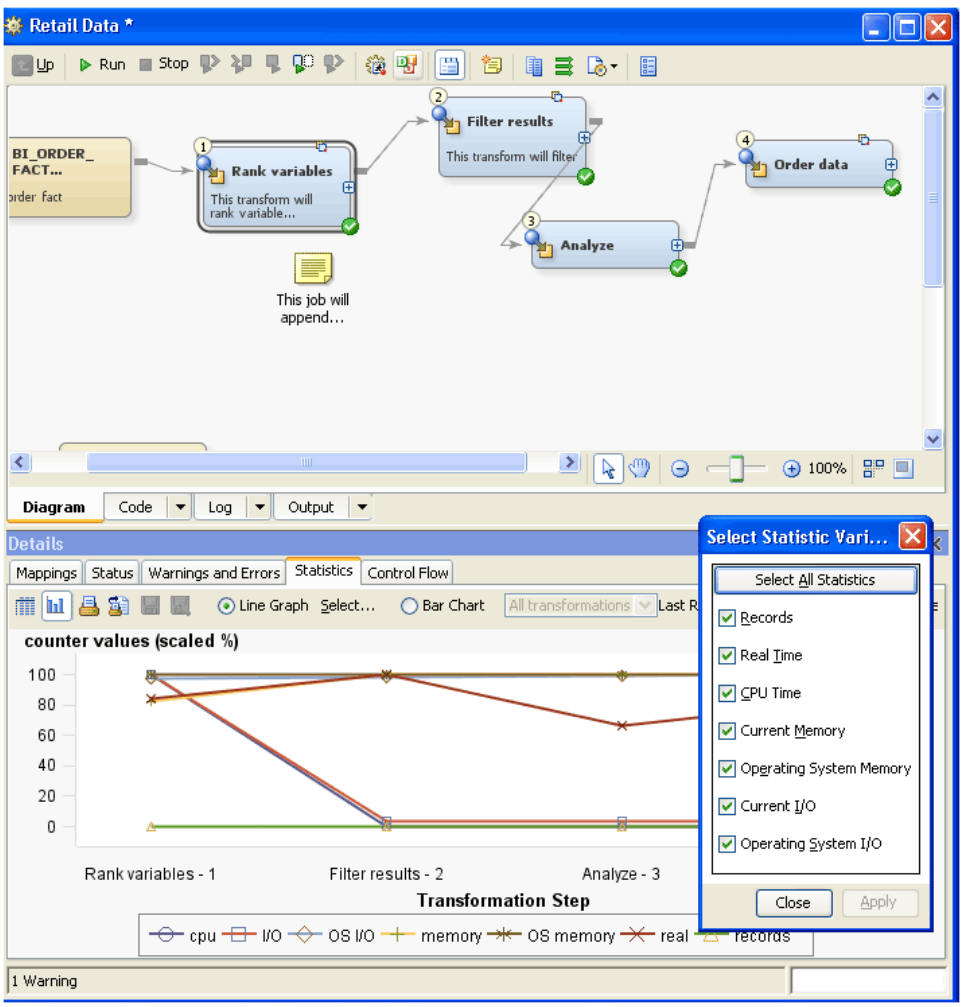
- Metadata is captured and documented throughout transformations and data integration processes, and is available for immediate reuse.
- Sophisticated metadata mapping technologies for quickly propagating column definitions from sources to targets, and for creating automated, intelligent table joins.
- New metadata search tool.
- Impact analysis for assessing the scope and impact of making changes to existing objects such as columns, tables and process jobs before they occur.
- Ability to determine the path, processes and transformations taken to produce the resulting information.
- Data lineage (reverse impact analysis), which is critical for both validating processes and building user confidence in data.
- Change analysis for metadata change discovery, comparison, analysis and selective propagation.
- Multiple-user collaboration support includes object check-in and check-out.
- Promotion and replication of metadata across development, test and production environments.
- Wizard-driven metadata import and export.
- Wizard for metadata column standardization.
- Metadata-driven deployment flexibility so that process jobs can be deployed for batch execution, as reusable stored processes or as Web services.



The Register Tables wizard makes it easy to access data from many different systems, as well as read and manage metadata from external sources.

Data cleansing and enrichment

- Data quality is embedded into batch, near-time and real-time processes.
- Data quality rules are callable through message queues, Web services and custom exits.
- Data cleansing is provided in native languages with specific language awareness and localizations for more than 20 regions worldwide.
- Data quality functions are available in both operational and reporting (transaction and batch) environments.
- Out-of-the-box standardization rules conform data to corporate standards, or you can build customized rules for special situations.
- Metadata built and shared across the entire process provides an accurate trail of actions applied to the cleansed data.
- Add value to existing data by generating and appending postal addresses, geo-coding, demographic data or facts from other sources of information.
- Data stewards can profile operational data and monitor ongoing data activities with an interactive GUI designed specifically for their needs.
- Simple process for institutionalizing data quality business rules. Apply basic or complex rules to validate data according to the specific business requirements of a particular process, project or organization. Rules may be applied in batch mode or as a real-time transaction cleansing process.
- Data quality monitoring enables you to continuously examine data in real time and over time to discover when quality falls below acceptable limits. Alerts can be issued when there is a need for corrective action.
- An interactive GUI enables you to profile operational data to identify incomplete, inaccurate or ambiguous data.
- Customizable and reusable data quality business rules that can be accessed directly within process job flows.



Often it is not enough to ensure that a job completes successfully as indicated by the green checkmarks in the lower right corner of each transformation step (top).

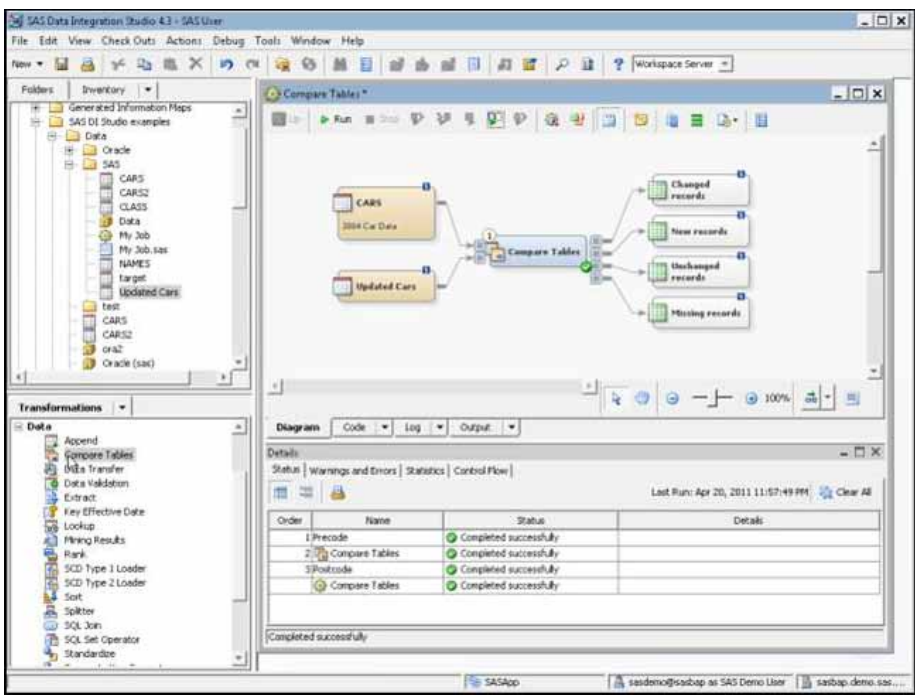
When working with large data flows, it is important to make sure that the job performs well. Saved metadata statistics are collected for each transformation step and can be viewed either graphically or in a tabular report (shown at bottom) to help tune performance.

Extraction, transformation and load (ETL) and extraction, load and transform (ELT)

- A powerful, yet easy-to-use transformation user interface that supports collaboration, reuse of processes and common metadata.
- Single or multiple-source data acquisition, transformation, cleansing and loading enable the easy creation of data warehouses, data marts, or BI and analytic data stores.
- Metadata is captured and documented throughout the data integration and transformation processes and is available for immediate reuse.
- Transformations can run on any platform with any data source.
- More than 300 predefined table and column-level transformations.
- Ready-to-use analytical transformations, including correlations and frequencies, distribution analysis and summary statistics.
- Transformation Generator wizard or Java plug-in design templates enable you to easily create reusable and repeatable transformations that are tracked and registered in metadata.
- Transformation processes are callable through custom exits, message queues and Web services, so they are reusable in many different projects and different technology environments.
- Transformations can be executed interactively and scheduled to run in batch at set times or based on events that trigger execution.
- Framework for publishing information to archives, a publishing channel, email or various message-queuing middleware.
- Easily refresh, append and update during loading.
- Optimize loading techniques with user-selectable options.
- Database-aware loading techniques include bulk-load facilities, index and key creation, and dropping and truncating of tables.
- Ability to easily design, create and load OLAP cubes.
- New transformations generate high-performance SAS code that is very efficient.
- New transformations include: Type 1 SCD support for merge and hash techniques, table differencing and enhancements for Type 2 SCD loaders.
- New Compare Tables transformation compares two data sources and detects changes in data.

Migration and synchronization

- Ability to migrate or synchronize data between database structures, enterprise applications, mainframe legacy files, text, XML, message queues and a host of other sources.
- Metadata-driven access to sources and targets.
- Extensive library of predefined transformations can be extended and shared with other integration processes.
- Embedded, reusable data quality business rules clean data as it is moved, synchronized or replicated.
- Recognizes changes to key fields and replicates/synchronizes changes across multiple databases.
- Optional, integrated scheduler allows changes made in one or more systems to be propagated to other systems on a scheduled basis.
- Delivers real-time data services for synchronization and migration projects.



The Compare Table transformation feature compares two data sources and detects changes in data.

Data federation

- Virtual access to database structures, enterprise applications, mainframe legacy files, text, XML, message queues and a host of other sources.
- Ability to join data across data sources for real-time access and analysis.
- Instant access to a real-time view of the data using the built-in data viewer.
- Query optimization is provided both automatically as part of DBMS requests, and manually within the advanced SQL editor, and can be used for both homogenous and heterogeneous data sources.

Master data management

- Enhanced metadata search features enable you to search by type, name, date or other keywords, subset by folders or other options, and save searches for future use.
- Support for semantic data descriptions of input and output data sources uniquely identify each instance of a business element (customer, product, account, etc.).
- Powerful transformation tools and embedded data quality processes ensure that master data is correct.
- Sophisticated fuzzy matching technology and innovative clustering methodologies enable you to validate and consolidate master records into identifiable data groups.
- Real-time data monitoring, dashboards and scorecards let you check and control data integrity over time.
- Can be used as a basis for transitioning to a full-fledged master data managemment offering.
- Data feeds can arrive in a single transaction or in hundreds of transactions at the same time.
- Data sets can be processed in a single pass of the source data.

Service-oriented architecture (SOA) Web services

- Open communication protocols for Windows and Java clients give developers access to SAS data integration and analytics from other programming languages, including Java, C++, VisualBasic.Net and more.
- SAS jobs and run streams can be called and executed remotely by developers without SAS programming knowledge.
- Access to static and streaming data for sending and receiving via Web services.
- Ability to leverage SAS data management capabilities from other business applications and systems via industry-standard Web services interface.
- Ability to expose standardized data quality and data mastering capabilities for real-time access in both internal and external business applications.

Message queuing

- Provides integration of asynchronous business processes via message-based connectivity.
- Interfaces to the leading message-queuing products, including Microsoft MSMQ, IBM WebSphere, Tibco Rendezvous and Java Message Service (JMS).
- Guaranteed message/transaction delivery reduces the cost of disruptions.
- Optimized access for each message-queue manager that is designed for minimal administrative effort.
- Event-based application integration so activities in one application automatically trigger actions in other applications.
- Dynamic, event-driven runstreams and alerts.
- Ability to send and receive messages between distributed and disparate systems.

SAS® Enterprise Data Integration Server System Requirements

To learn more about SAS Enterprise Data Integration Server system requirements, download white papers, view screenshots and see other related material, please visit www.sas.com/ediserver.

SAS Institute Inc. World Headquarters +1 919 677 8000
To contact your local SAS office, please visit: **www.sas.com/offices**