# SAS® Data Integration for Midsize Business

Base your decisions on complete, consistent data

## What does SAS® Data Integration for Midsize Business do?

SAS Data Integration for Midsize Business is a powerful, configurable and affordable solution that can meet a wide variety of data integration requirements, from small departmental projects to strategic business initiatives. With it, you can access virtually all data sources to extract, cleanse, transform, conform, aggregate, load and manage data. It also supports data warehousing, migration, synchronization and federation initiatives, as well as batch-oriented and real-time master data management solutions. Use the solution to create reusable data integration services that support service-oriented architectures and data governance, which can be used in other projects at a later date.

## Why is SAS® Data Integration for Midsize Business important?

It enables midsize businesses to efficiently manage simple to complex data integration projects in a timely, cost-effective manner so you can meet the demands of information consumers.

## For whom is SAS® Data Integration for Midsize Business intended?

It is designed for small and midsize businesses in all industry sectors that need to implement one or more data integration projects, deal with changing business landscapes and meet regulatory requirements or implement data governance.

Business users depend on data to be trustworthy, complete and readily available when they need it. But with data spread across disparate systems – and rapidly increasing data volumes – data integration is no easy task.

SAS Data Integration for Midsize Business is a simple, flexible solution that addresses the challenges of data integration projects at small and midsize businesses. It ensures data credibility and consistency – enabling organizations to easily manage all their data integration projects while reducing costs and increasing overall productivity.

The solution includes a common repository for centralized storage, management and reuse. This reduces both development and maintenance time, and minimizes pressure on limited IT resources. Users can work collaboratively in groups or teams, and a common set of tools shortens the learning cycle and simplifies maintenance. Reusable templates make it quick and easy to provide role-based authorizations and administrative privileges at the user, departmental or organizational level.

## Key Benefits

- **Always access the data you need.** From legacy systems to Excel or the latest ERP applications, data from virtually any hardware platform or operating system can be accessed and processed using SAS/ACCESS® engines. Add new source systems easily and manage security centrally. This saves time, shortens learning curves and gives you the information and confidence you need to make data-driven decisions.

- **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 information. Powerful data lineage tools enable users to see where data originated and how it has been transformed. The solution's data profiling tools provide insights into the quality of processes and source systems so business users know they are using the best data possible.

- **Reduce costs by eliminating overlapping, redundant tools and complex system architectures.** This flexible, reliable solution supports both the operational and decision support needs of small and midsize organizations. It minimizes the time and effort associated with piecemeal approaches that entail linking and managing technologies from different vendors – to ensure data credibility, reduce risk and speed results.

## Solution Overview

SAS offers the only data integration solution built from the ground up to meet the data integration needs for small and midsize businesses. The solution provides a collaborative design environment that promotes object reuse and sharing. It also offers administrative controls and a wizard-driven design process workflow. You can access data from virtually any system in any form, transform and cleanse the data in near-real time, and handle data migration, synchronization and federation projects within an environment that is easy to use, deploy and maintain. SAS Data Integration for Midsize Business has all the capabilities and flexibility needed to meet growing data and changing business requirements.

### Interactive data integration development environment

The GUI provides users with an interactive and intuitive set of configurable windows for building, governing and maintaining reusable data integration services. Wizards make it easy for new users to quickly add value to the team.

### Connectivity across the organization

Comprehensive connectivity is the foundation of a complete data integration solution. SAS Data Integration for Midsize Business provides connectivity to all major and most other data sources, operating systems and hardware environments using both native access and open standards.

### Data profiling

SAS Data Integration for Midsize Business includes powerful data profiling capabilities that provide the ability to analyze and assess the quality of data within single or across multiple source systems. You can quickly identify where potential problems exist and how much effort is required to rectify them. This enables you to focus on the root cause of data quality issues, and more accurately plan and execute your data integration projects.

### Extract, transform and load (ETL) and extract, load and transform (ELT)

Loading data warehouses and data marts, building analytical marts for special projects, and creating extract files for reporting and analysis applications can consume limited IT resources, leaving them with little or no time for strategic initiatives.

SAS Data Integration for Midsize Business includes an intuitive point-and-click design editor window. It helps users to easily build logical process workflows, quickly identify the input and output data stores, and create business rules in metadata – enabling rapid generation of data warehouses, data marts and data streams. Users can also choose to have transformations and processes take place directly inside a connected database or storage system instead of moving the data. Referred to as ELT, push-down or in-database processing, this reduces unnecessary data movement to substantially speed up overall processing times, allowing you to be faster and more productive.

### 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
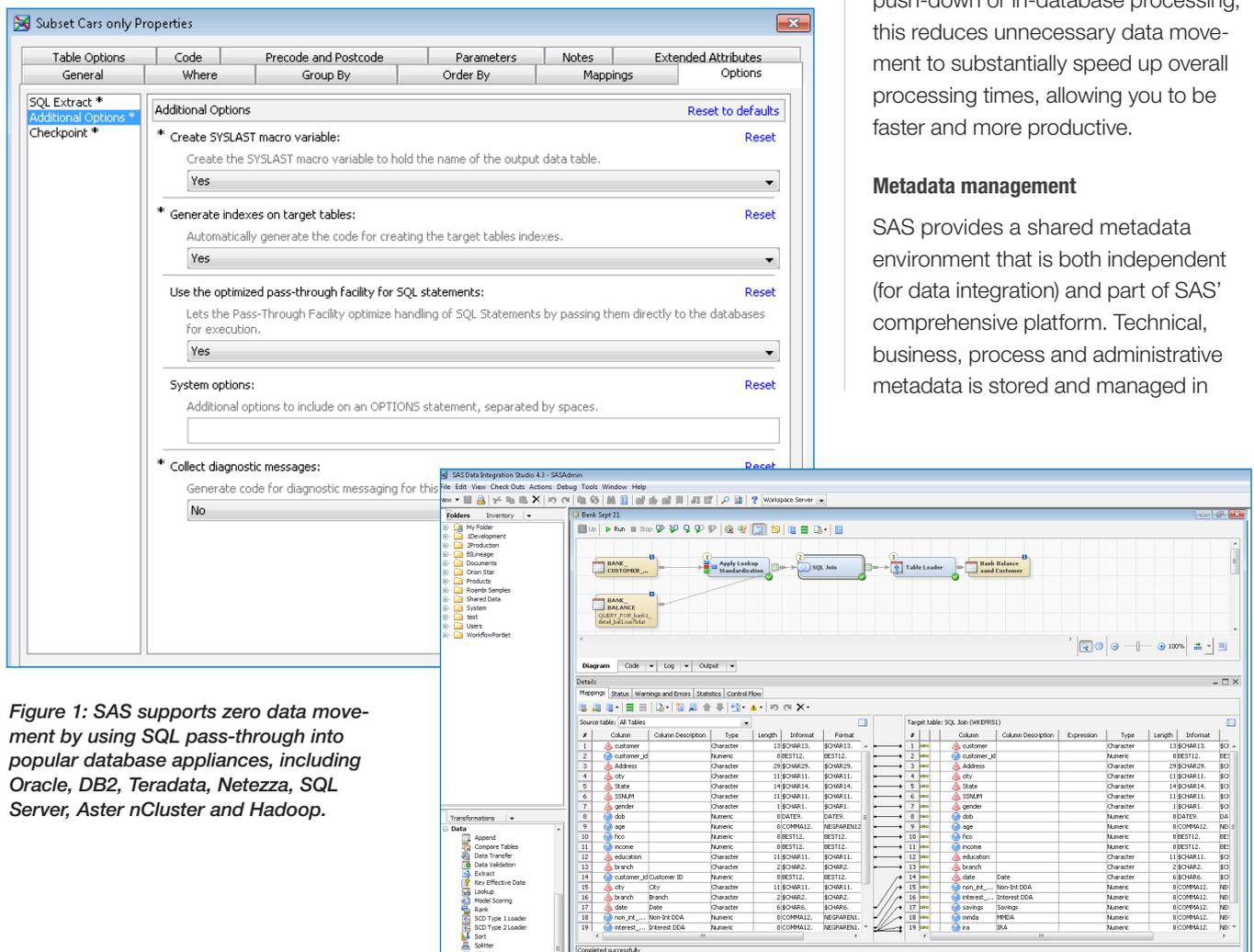


*Figure 1: SAS supports zero data movement by using SQL pass-through into popular database appliances, including Oracle, DB2, Teradata, Netezza, SQL Server, Aster nCluster and Hadoop.*



*Figure 2: Source-to-target data mapping.*

a way that facilitates reuse of existing table definitions, business rules and more. Shared metadata provides a consistent definition across data sources to speed integration projects, simplify design and reduce maintenance costs.

**Migration and synchronization**

Moving data from system to system is a constant activity in most organizations. For example, 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.

With SAS Data Integration for Midsize Business, you can migrate, synchronize and replicate data across different operational systems and data sources. The 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**

Having fast access to current operational data is critical for your organization's success. Your business users need to be able to independently access and analyze data across multiple sources so they can make timely decisions. SAS Data Integration for Midsize Business lets users query and use data

## Key Features

**Interactive data integration development environment**

- Wizards for accessing source systems, creating target structures, importing and exporting metadata, and building and executing data access, transformation and load process flows.
- Dedicated GUI to profile data in order to identify and rectify source system issues, while retaining the business rules for later use in ETL processes.
- Multiuser design environment supports collaboration on large projects.
- Ability to distribute data integration tasks to nearly any platform and to connect virtually any source or target data store.
- Integrated workflow scheduling, automatic load balancing and grid computing support.
- Design-time and run-time visualization and monitoring.
- Job statistics and logging.
- Enhanced GUI-based debugging.
- Push-down indicators that show when transformation processing is being passed to and performed by a connected database/data warehouse/data storage system.
- Checkpoint/restart so that job and data load processes are tracked, enabling interrupted processes to be restarted at the point where they were interrupted.
- Powerful data profiling capabilities.
- Ability to import existing SAS programs and convert them to GUI-based jobs by automatically defining metadata attributes and creating the associated process job flows.

**System connectivity**

- Provides connectivity to all major and most other data sources, operating systems and hardware environments using both native access and open standards.
- Supports unstructured and semistructured data to parse and process files.
- Gives access to static and streaming data for sending and receiving through Web services.

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

- Transformation library with more than 300 predefined table and column-level transformations.
- Transformation generator wizard for creating reusable and repeatable transformations that are tracked and registered in metadata.
- Framework for publishing information to archives, email or various message queuing middleware.
- Metadata is captured and documented throughout transformation and data integration processes, and is available for immediate reuse.
- Flexible deployment: Transformations can run on all supported platforms with all data sources. Transformations and data integration processes can be deployed easily as embedded business logic for use by other applications.

**Migration and synchronizations**

- Metadata-driven access to sources and targets.
- Extensive library of predefined transformations for migration and synchronization that can be extended and shared with other integration processes.
- Embedded, reusable data quality business rules can clean data as it is moved.
- Ability to migrate or synchronize data between database structures, software applications, legacy files, text, XML, message queues and a host of other sources.
- Change data capture (CDC) capabilities recognize changes to key fields and can replicate/synchronize changes across multiple, heterogeneous databases.

**Data federation**

- Virtual access to database structures, enterprise applications, legacy files, text, XML, message queues and a host of other sources.
- Ability to join data across data sources for access and analysis.
- Instant access to a real-time view of the data using the built-in data viewer.

across multiple systems without physically reconciling or moving source data. By avoiding unnecessary data replication and movement, you can quickly and cost-effectively deliver current data that is consistent and accurate.

**Integrated support services**

The features and capabilities of SAS Data Integration for Midsize Business can be used in a modular manner with an array of supporting data integration services, including scheduling, connectivity, administration, data enrichment, data movement, data parsing, data cleansing, data auditing, data profiling and others. All of these supporting services utilize the integrated SAS metadata environment and administrative functions. With the help of easy-to-use wizards, users can easily create their own data integration services that can be called in real time, near-real time and batch modes.

**SAS® Data Integration for Midsize Business System Requirements**

To learn more about SAS Data Integration for Midsize Business system requirements, download white papers, view screenshots and see other related material, please visit sas.com/di-smb.

## Key Features (continued)

- 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.
- Optimized loaders are designed to use various load techniques and bulk-loading utilities.
- A specially designed loader provides additional optimized support for Teradata.

### Metadata management
- Sophisticated metadata mapping technologies for quickly propagating column definitions from sources to targets, and for creating automated intelligent table joins.
- Impact analysis for assessing the scope and impact of making changes to existing objects, such as columns, tables and process jobs.
- Data lineage within the SAS environment validates processes that are working correctly and builds user confidence in data.
- Wizard-driven metadata importing and exporting.
- Metadata-driven deployment flexibility: Process jobs can be deployed for batch execution, as reusable stored processes and Web services.
- Metadata reports provide an easy-to-understand view of the metadata in HTML format for users to view the content.

### Integrated support services
- Features and capabilities provided by SAS Data Integration for Midsize Business can be used modularly with an array of supporting data integration services.
- Easy-to-use wizards for creating services including scheduling, connectivity, administration, data enrichment, data movement, grid support, data parsing, data cleansing, data auditing, data profiling, data lineage and data transformation.
- Integrated SAS metadata environment and administrative functions.

**sas**
THE POWER TO KNOW®