



**The SAS[®] Enterprise Intelligence Platform:
SAS[®] Data Integration**

Three steps to faster ROI from your information assets

SAS[®]9

Table of Contents

Executive summary	1
Key challenges to effective enterprise data stewardship	2
Toward a transformation in data integration strategy	4
Optimize your existing infrastructure with the integrated	
SAS Enterprise Intelligence Platform	4
Components of the SAS Enterprise Intelligence Platform	5
SAS Data Integration	6
Components of SAS Data Integration	6
Rapid development.....	7
Efficient processing	7
Greater management control	8
Key benefits of SAS Data Integration	9
Faster, less expensive development of data integration processes	9
More efficient processing for lower cost of ownership	9
Greater control and accountability for more effective data stewardship	10
Increased value from legacy systems, hardware and ERP investments	10
Summary	11
From SAS, the first name in enterprise intelligence solutions	11

Executive summary

Accurate business decisions depend on clean, accurate data. But identifying, validating, collecting, manipulating and delivering the right data can be an expensive and unreliable process with numerous challenges.

Many companies operate in silos, where information is scattered throughout the enterprise and stored in inconsistent formats. The proliferation of heterogeneous platforms has made it hard to identify inconsistent and duplicate data during data integration, which contributes to the rising cost of data stewardship. Data representations can vary among systems, and it is hard to manage the metadata that defines how data elements were derived and should be managed. Even with massive investments in distributed computing power, IT groups are still struggling to run data integration processes within allotted batch windows.

As a result, many IT organizations continue to search for systematic, predictable methods to assure data quality and timeliness — under the constraints of shrinking budgets and reduced headcount.

The SAS Enterprise Intelligence Platform provides an integrated platform that overcomes these challenges and delivers enterprise intelligence while extending the value of existing IT architectures.

The **SAS Data Integration** component of the SAS Enterprise Intelligence Platform delivers prebuilt, high-performance capabilities for data connectivity, data quality, ETL (extract, transform and load), data migration, data synchronization and data federation, and improves the return on investment of your information assets in three key ways:

- **Rapid development.** Powerful data integration process development tools, with drag-and-drop ‘wizards,’ enable rapid creation of customizable data integration strategies without requiring in-depth programming expertise or maintenance of thousands of lines of custom code.
- **Efficient processing.** Data integration efficiency is maximized through integrated data quality routines, one-step processing of multiple outputs, data integration processing on the platform where the data resides, automated workflow scheduling with load balancing and more.
- **Greater management control.** SAS Data Integration introduces a new level of accountability and control over all aspects of the data integration process — from auto-generated audit trails of process changes over the project life cycle, to easy-to-use metadata management and a single administrative window into information systems.

Read on for an overview of the SAS Enterprise Intelligence Platform and the key offerings and benefits of the SAS Data Integration component, which provides the essential data foundation for a unified intelligence architecture.

SAS Data Integration can read and extract all data formats from heterogeneous platforms without relying solely on SQL.

It offers prebuilt, high-performance capabilities for data connectivity, data quality, ETL, data migration, data synchronization and data federation.

Key challenges to effective enterprise data stewardship

It is a trying time for enterprise IT managers, who are under pressure to do more with less — to reduce costs and extend the value of existing investments while increasing strategic contributions to the organization. Without the ability to provide quality data across the enterprise, IT is at risk of losing credibility within the organization or not meeting internal service level agreements.

Collecting and cleansing the right data to support business intelligence and analytic requirements often is an expensive and unreliable process filled with challenges.

Still, most IT managers will admit that collecting and cleansing the right data to support business intelligence and analytic requirements is often an expensive and unreliable process filled with challenges:

Data integration process development is too complex, too slow and too costly.

As businesses have grown organically and through mergers and acquisitions, many have evolved information silos in which operational data has been captured in a wide variety of sources and formats, such as ERP systems, RDBMS (relational database management systems), flat files and Web logs. Individual business units often purchased applications and storage independent of a corporate IT plan. A side effect of this is that it is now difficult for IT departments to locate, identify and select the data that decision makers need. As a result, data integration processes can be slow to develop and expensive to maintain.

In an environment of disparate applications, intelligence exists but it is hidden in silo systems that share their information in limited ways, if at all.

IT is often forced to manually write excessive code to carry out seemingly standard transformations required in the data integration process. This code often is created in an ad hoc way by individual developers — an inefficient and error-prone approach. There's often no effective way to share and maintain these custom-coded transformations. It is difficult to leverage previous custom work, and there are usually few or no audit trails except the developer's recollections.

Data integration processes take too long to run and often leave data quality problems unresolved.

Many organizations are investing heavily in additional compute power and resources to meet demands for high-volume processing and tight turnarounds. Yet despite all of the hardware investments, many distributed CPUs are under-utilized. It is still a struggle to get data integration processes to finish during the batch window when key servers are busy. Manual scheduling outside of that window uses resources inefficiently and ultimately increases support costs and execution cycles. A better way to manage and balance the computing power already in place is needed.

The quest for efficiency is hindered by processes that are inherently inefficient. For example, SQL-only data integration tools can result in unacceptable extraction times when large numbers of tables are joined. SQL-based data integration tools have difficulty creating multiple output tables in one pass of input data, forcing expensive multiple-pass processes. IT data centers may find they are unable to load all the required data in the overnight batch window — especially if an RDBMS is asked to load and index large volumes of data.

Moreover, many data integration processes force IT to move data from source platforms to another location in order to transform it, and in doing so, they compromise data security, cause undesirable data duplication and increase hardware costs just to compensate for the limitations of the tool. In addition, SQL alone often cannot satisfy requirements to read and extract core data in all situations. For example, SQL falls short when you're dealing with variable record lengths, repeating groups and unconventional formats and data stores — or when extracting data from systems with layers of cryptic metadata, such as ERP and front-office CRM systems.

Sometimes this process involves moving the data from operational systems onto new platforms so that it can be used by analytical and business intelligence applications — a workaround that burdens the network with excessive traffic, redundant data and security risks.

Consider also the inherent inefficiency of the typical multivendor data integration framework. In “best-of-breed” strategies, multiple tools are often required to extract data from various databases, ERP systems and other sources to bring together a single version of the truth. Managing all these separate tools further strains IT resources and slows productivity.

Multiple tools make it difficult to identify inconsistent and duplicate data during data integration, which affects the rising and ongoing cost of data stewardship. IT departments continue to search for systematic, predictable methods to assure data quality, though it is often unclear how much effort will be required to integrate and cleanse data to produce information the business can trust.

It's hard to manage and track data integration processes and changes.

If multiple developers are customizing a data integration process, how are changes managed, controlled and tracked? In large projects, it can be time consuming to document every detail of who did what, when things changed and where it fits into the big picture. Typically, this documentation is lacking, resulting in delays as developers search for things that could be found easily if appropriate documentation existed and try to work out how data elements have been derived, who made which changes and why.

Without full confidence in the metadata that defines data, you can't have full confidence in the analytic reports that rely on that data.

All of these problems are compounded when there is no mechanism to automatically create and manage metadata. Documentation about metadata — such as what the data element is, where it originated and how it was transformed to its current form — is often not available or is out-of-date. Or, it might be stored in multiple systems with multiple representations, making those systems nearly impossible to integrate. The data element “apples” on one system might actually be coded as “NotOranges” on another platform or as “apples_1” or “apps” — and may or may not be derived in the same way. It can be difficult or impossible to accurately consolidate information from different systems based on exact value matches, such as Table 1 (name) = Table 2 (name). Data consolidation is then incomplete and inaccurate, and the business user receives poor quality information for decision making.

Furthermore, in the typical distributed environment, the IT group has to manage licenses with one administrative tool, manage user and group definitions with another and manage metadata with yet another tool — all of which have a different look and feel. This is a tremendous waste of time, money and resources.

If these challenges seem familiar, you are in good company. These scenarios are prevalent throughout all industries and information architectures. The main challenge now is to implement a high-quality data integration strategy that will meet the organization's needs going forward.

Toward a transformation in data integration strategy

In devising data integration strategies, IT managers have to resolve historical problems related to data quality and consistency, and continue to improve end-user confidence in the applications and systems that rely on that data. To overcome those obstacles and deliver reliable information to decision makers, companies need a data integration platform that achieves three key benefits:

- **Accelerates development** with development tools that make it easy to create powerful, customized data integration strategies, without requiring in-depth programming expertise or maintenance of thousands of lines of custom code.
- **Maximizes data integration efficiency** by synthesizing corporate data from many platforms, integrating data quality processes, and scheduling and load balancing data integration processes to fit into allowed time windows.
- **Provides better accountability and control** making it easier to administer data integration processes, track process changes and manage metadata.

Successful business intelligence and analytic applications are only as agile and dependable as the data they digest. With competitive intensity at an all-time high, it is time to reconsider the processes that underpin enterprise intelligence, beginning with the essential data foundation.

Your existing infrastructure is core.

The SAS Enterprise Intelligence Platform brings it all together.

Optimize your existing infrastructure with the integrated SAS Enterprise Intelligence Platform

SAS has developed an integrated platform for delivering enterprise intelligence. This platform, which we call the SAS Enterprise Intelligence Platform, optimally integrates individual technology components within your existing infrastructure into a single, unified system. The result is an information flow that transcends organizational silos, diverse computing platforms and niche tools — and delivers access to the insights that drive value for your organization.

The SAS Enterprise Intelligence Platform extends the value of your existing systems while setting the stage for new levels of enterprisewide intelligence not previously possible. In the process, you'll jump-start the evolution of the IT organization from cost center to strategic partner.

Components of the SAS Enterprise Intelligence Platform

The SAS Enterprise Intelligence Platform allows you to optimally configure the technology components within your architecture to deliver intelligence throughout the organization at the lowest total cost of ownership and with the fastest time to intelligence.

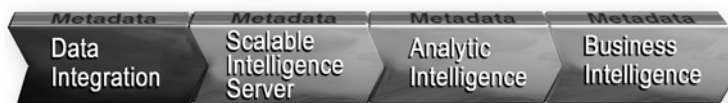
- **SAS Data Integration** provides prebuilt, high-performance capabilities for data connectivity, data quality, ETL (extract, transform and load), data migration, data synchronization and data federation.
- **SAS Scalable Intelligence Server** is a dedicated solution that efficiently stores and disseminates information for business intelligence and analytic requirements, offering relational and OLAP storage options from the same foundational inputs.
- **SAS Analytic Intelligence** is an integrated environment for predictive and descriptive modeling, forecasting, optimization, simulation, experimental design and more. SAS Analytic Intelligence leverages existing data and infrastructure to support effective decision making and integration into business intelligence environments.
- **SAS Business Intelligence** delivers a set of BI capabilities that enable different types of users to surface meaningful intelligence from consistent, companywide data.

As part of an integrated platform, all components and services are managed from a single point, reducing the administrative effort for maintenance of applications, users and security. Data consistency is assured because metadata is stored in a single metadata repository and is shared across all SAS technologies and solutions. Supporting a wide range of open standards, the SAS Enterprise Intelligence Platform has been designed to facilitate integration with existing IT infrastructures.

When integrated into a single, cohesive technology platform, these components help you optimize your current intelligence environment and better align IT with the strategic objectives of the organization.

SAS Data Integration

A high-performance data integration platform



SAS Data Integration synthesizes corporate data from disparate silos of information in a timely, cost-effective manner. SAS Data Integration is capable of reading and extracting all data formats from heterogeneous platforms, without relying solely on SQL, and offers prebuilt SAS capabilities for data manipulations, quality and analysis. SAS Data Integration improves the return on investment of information assets in three key ways:

- **Rapid development.** Development tools, with drag-and-drop 'wizards,' enable rapid creation of powerful, customizable data integration strategies without requiring in-depth programming expertise or maintenance of thousands of lines of code.
- **Efficient processing.** Data integration efficiency is maximized through integrated data quality routines, one-step processing of multiple outputs, data integration processing on the platform where the data resides, automated workflow scheduling with load balancing and more.
- **Greater management control.** SAS Data Integration introduces a new level of accountability and control over all aspects of the data integration process — from auto-generated audit trails of process changes over the project lifecycle, to easy-to-use metadata management and a single administrative interface into information systems.

Components of SAS Data Integration

SAS Data Integration provides the modules and tools required to achieve rapid development, efficient processing and greater management control:

- An interactive data integration process development environment.
- Metadata management facilities.
- Read/write capabilities to third-party databases.
- Network connectivity facilities.
- Deployment and management facilities.
- Data quality and cleansing capabilities.

Rapid development

The interactive data integration process development environment provides patent-pending ‘wizards’ to simplify data integration process design, even for complex processes that involve many data and analytical transformations and draw from many data sources. The wizards automatically connect to the selected data source and populate the metadata repository with relevant information, such as descriptions and keys and all relationships.

A drag-and-drop user interface on the design tool gives the data integration project developer a panoramic view of diverse data sources and a network view, making it easy to design and manage data integration processes. Prepackaged ERP and front-office data integration templates support frequently occurring business intelligence scenarios.

Developers now have a single point of control for managing the entire data integration process. The design tool can direct any portion of the data integration process to run on any platform and/or against any database natively, including third-party relational databases.

The SAS Data Integration transformation engine provides thousands of built-in transformation functions that run on all major operating systems, including MVS, UNIX and Windows — minimizing large data transfers and optimizing resource costs. Transformation ‘wizards’ with drag-and-drop interfaces remove the burden of managing thousands of lines of custom code, and yet prebuilt transformations can be customized easily. Whether customized or out-of-the-box, all transformations are tracked and registered via central metadata, which streamlines project management.

In an environment of disparate applications, intelligence exists but it is hidden in silo systems that share their information in limited ways, if at all.

Efficient processing

Fully integrated data quality routines accelerate data cleansing — deduplicating and cleansing data as an integral part of the data integration process. This integrated approach reduces ongoing cost of data stewardship while increasing data accuracy. Data quality algorithms run on the industry’s widest range of platforms, from Windows to mainframes. Developers can define custom corporate data interpretation standards using an editable knowledge base and custom interfaces.

Processing efficiencies provide maximum flexibility while reducing the time and IT resources required to run data integration processes. A set of open and native multiprocessing data extraction and transformation engines can be deployed on a wide variety of platforms and can read every file and RDBMS format encountered in practice. The optimized storage mechanism loads and indexes faster than traditional RDBMS. Multiple inputs and outputs can be processed in a single step, without the need to reread input sources.

SAS provides the only data integration solution with integrated data cleansing routines that are both customizable and usable out-of-the-box.

Integrated workflow scheduling and load balancing allow easy and efficient execution of data integration processes. Load balancing makes execution times predictable by automatically redirecting jobs and processes to the resources with the least load — allowing for process inter-dependencies and enterprise or departmental resource constraints. This capability optimizes hardware utilization, reduces processing cycle time and cost, and increases overall productivity.

Integration with market-leading message-oriented middleware (MOM) enables near real-time data integration processing. Message queues provide a loose coupling between applications, using an asynchronous, disconnected, guaranteed-deliver, store-and-forward architecture. The use of MOM lets diverse applications communicate asynchronously without a direct communication link between applications, and guarantees that messages will be delivered to the destination. In a world of distributed systems that cannot easily communicate with each other, SAS' integration with MOM helps IT keep data integration processes in step with the demand for timely enterprise information.

Greater management control

Integrated metadata management and standards enforcement facilitates the creation, storage and exchange of metadata both within and outside the enterprise. This completely open enterprise metadata structure retains and tracks technical, business and administrative metadata and aids in developing data integration processes. This metadata repository documents data integration paths from end to end, eliminating guesswork and tedious research to find out how data elements are derived.

An integrated management console provides powerful tools for deploying and maintaining information assets across the enterprise.

Metadata information is gathered by the data integration process itself, or automatically from other metadata tools and databases. Navigational tools clarify how the data was derived — from its source, how it was transformed, by whom, and to its destination(s). This level of detail enables instant reconciliation against any other source of information.

With compliance to Common Warehouse Metamodel (CWM) and Meta Integration Technology, Inc. (MITI) standards, SAS Data Integration processes can interchange metadata with other standards-compliant metadata sources.

An integrated management console provides powerful tools for deploying and maintaining information assets across the enterprise (both SAS and third-party resources), enforcing standards and maintaining authorizations and systems of record for information. By using a single administrative interface, this management console reduces administrator training, simplifies administrative tasks and enables repeatable processes that uphold best practices and minimize manual tasks.

The integrated data integration change management system enables implementation teams to synchronize work and collaborate on large data integration projects. The system logs a full history of changes made by pre-authorized developers throughout the project life cycle. Full check-in/check-out on data integration process and individual metadata elements results in automatic, up-to-date documentation and easy-to-use metadata management. The results: development time is dramatically reduced, an accurate audit trail is available to create repeatable processes and ongoing maintenance is simplified.

Your data integration processes form the foundation of enterprise intelligence. The quality of data integration determines data quality and completeness, the ease with which data can be applied in analytical applications and adaptability to new data sources later. SAS Data Integration will help you build a solid data foundation on which decision makers can depend.

Key benefits of SAS Data Integration

Faster, less expensive development of data integration processes

SAS Data Integration provides automatic, up-to-date documentation and easy-to-use metadata management, so development and maintenance take less time. The solution provides full support for SQL while offering a wide range of pre-defined transformations and a number of out-of-the-box access engines for industry-leading RDBMS and ERP systems. These capabilities reduce the need for custom integration solutions and ensure that IT can access all available enterprise data.

By addressing the full spectrum of ERP and legacy data formats and platforms, SAS Data Integration makes all metadata available from an integrated environment, which further lowers the cost of data integration software tools and training. SAS' data integration platform can access and extract from RDBMS, flat files, Web logs and legacy sources as well as navigate, identify and extract data from systems that have an additional layer of metadata, such as ERP systems and front-office CRM systems.

More efficient processing for lower cost of ownership

SAS' integrated data quality routines resolve data quality and consistency problems, thereby reducing the ongoing costs of data stewardship. These data quality routines can be used out-of-the-box or customized to meet the needs of the organization.

SAS Data Integration also allows the data integration process to take place on the platform where the data resides, a powerful capability that eases network traffic because data is not moved from its native platform. In this way, the security and integrity of the original data stores are maintained and additional data duplication and hardware requirements are reduced.

The powerful SAS transformation language minimizes hardware investments needed to optimize the data integration process. This transformation language provides robust and flexible capabilities within multivendor architecture environments, making it easy to add new data sources, for example, after corporate mergers and acquisitions.

Greater control and accountability for more effective data stewardship

A unified management console interface coupled with a detailed audit trail of process changes simplify ongoing maintenance, reduce administrator training and make it possible to create repeatable processes that uphold best practices and corporate policies. A unified repository of end-to-end metadata makes it easy to discern how data elements were derived, and enables sharing of data and applications among business intelligence and analytic intelligence systems from SAS or third parties.

Increased value from legacy systems, hardware and ERP investments

SAS Data Integration provides a single point of control for efficiently managing data integration processes and building a data warehouse across all legacy data sources. Using this platform, there is no need to retire legacy systems or invest in expensive, hard-to-maintain custom interfaces simply to generate reports from them.

For example, many organizations that originally implemented ERP systems to improve operational efficiency have incurred significant costs in attempting to adapt those systems to meet demands for business intelligence and analytics. The results often are inefficiently generated, inadequate enterprise intelligence and reduced ROI on ERP investments. SAS Data Integration overcomes the obstacles of extracting and integrating data from ERP systems, allowing you to separate investments in ERP from analytical and business intelligence systems and maximize the ROI of each.

SAS Data Integration provides integrated scheduling and load sharing as well, so all available IT hardware can be utilized during the data integration process, ensuring that IT maximizes its use of data center assets.

Summary

As a component of the SAS Enterprise Intelligence Platform, SAS Data Integration forms a solid data foundation on which decision makers can depend. A data integration solution engineered with SAS increases ROI from existing IT investments through:

- **Rapid development** using automated, wizard-based development tools with prebuilt SAS capabilities simplify data integration processes that may include thousands of data manipulations, quality routines and metadata analysis functions.
- **Efficient processing** with multithreaded/multiprocessing data extraction and transformation engines that run simultaneously on multiple computing platforms, integrated data quality routines that can be used out-of-the-box or customized, automated workflow scheduling with load balancing and more.
- **Greater management control** over all aspects of the data integration process — from auto-generated audit trails of process changes over the project lifecycle, to easy-to-use metadata management and a single administrative interface into information systems.

Processing efficiencies consume fewer resources and data manipulation capabilities meld disparate silos of information into an integrated platform for enterprisewide intelligence.

The cumulative effect of these gains extends the value of existing IT infrastructures — from processing efficiencies that consume fewer resources and forestall the need to acquire more hardware, to data manipulation capabilities that meld disparate silos of information into a future-ready platform for enterprisewide intelligence.

From SAS, the first name in enterprise intelligence solutions

For nearly three decades, SAS has been helping customers realize the full potential of enterprisewide intelligence. Through listening to our 40,000 customers across every industry — 96 of the top 100 companies on the FORTUNE Global 500® — SAS has developed an integrated platform for delivering high-value enterprise intelligence.

This platform, which we call the SAS Enterprise Intelligence Platform, optimally integrates individual technology components within your existing infrastructure. SAS is the only vendor that completely integrates leading data integration, storage, analytics and traditional business intelligence applications to create intelligence from massive amounts of data. SAS Data Integration provides high-performance data integration for the SAS Enterprise Intelligence Platform.

Let us show you how the SAS Enterprise Intelligence platform can deliver immediate and sustainable rewards for your organization. Whatever your existing data infrastructure, SAS can transform it into a foundation for enterprisewide intelligence and sustainable profitability in an increasingly uncertain world. For more information about the SAS Enterprise Intelligence Platform and the specific SAS technologies that support it, visit us at www.sas.com.



World Headquarters
and SAS Americas
SAS Campus Drive
Cary, NC 27513 USA
Tel: (1) 919 677 8000
Fax: (1) 919 677 4444
U.S. & Canada sales:
(1) 800 727 0025

SAS International
PO Box 10 53 40
Neuenheimer Landsr. 28-30
D-69043 Heidelberg, Germany
Tel: (49) 6221 4160
Fax: (49) 6221 474850

www.sas.com