



SAS® Data Integration Server

A complete solution designed to meet your data integration needs

What does SAS® Data Integration Server do?

SAS 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 reusable data integration services in support of service-oriented architectures and data governance.

Why is SAS® Data Integration Server important?

It enables organizations to efficiently manage simple to complex data integration projects in a timely, cost-effective manner and to meet the needs of information consumers.

For whom is SAS® 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.

SAS Data Integration Server provides a comprehensive solution to handle the challenges of distributed and rapidly increasing data volumes, inconsistently defined data across disparate IT systems and the high expectations of data consumers who depend on data to be correct, complete and available when they need it. SAS Data Integration Server does this in a timely, cost-effective manner, ensures credibility and consistency, and gives organizations the ability to efficiently manage data integration projects, increasing overall productivity and reducing the total cost of ownership.

Key benefits

- **Leverage/reuse work by others.** A common repository enables the centralized storage, management and reuse of work created by all users based on their authorizations so it is very easy to create, share and use jobs developed by others. This reduces both development and maintenance time.
- **Work collaboratively in groups or teams.** Leveraging the common repository, multiuser development capabilities are supported. In addition, using a common set of tools among developers shortens learning curves and makes maintenance easier. The advanced GUI environment and centralized metadata repository capture and manage the work so that it is easy to see and understand, and is easier for new team members to become productive right away.
- **Remove technology obstacles that endanger project deadlines.** Easy to learn and work in, the advanced GUI environment provides developers with a standard interface for building and documenting their work. Manual coding, though available when wanted, is minimized. Users are provided with hundreds of prebuilt capabilities as well as the ability to create their own new data integration services, all from within the SAS environment.
- **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 and processed using SAS/ACCESS® engines. New source systems can easily be added and security managed centrally. This saves time, shortens learning curves and gives decision makers the information they really need.
- **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 developers and users to see where data originated and how it has been transformed. In addition, data quality profiling tools, featuring DataFlux technology, provide insights into the quality of processes and source systems so users are assured they are using the best data possible.
- **Manage security and administration at all levels.** With its centralized common metadata repository, SAS enables security and a single point of administration for all levels. Reusable templates make it quick and easy to provide role-based authorizations and administrative privileges at the user, departmental or organizational level.



- **Meet time constraints in ever decreasing windows of availability.** SAS processes data fast! Organizations can take advantage of the grid-enabled, load-balanced, multithreaded parallel processing architecture that can quickly process, transform and move data between different platforms and systems.
- **Eliminate overlapping, redundant tools and systems with one solution.** SAS Data Integration Server is a flexible, reliable and complete solution that strategically supports all the needs of an organization, including both operational and decision support/business analytic solutions. 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 your data integration needs. Instead of linking and managing technologies from different vendors, SAS 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) provides developers with an interactive and intuitive set of configurable windows for building, governing and maintaining reusable data integration services as part of the data integration development processes. Wizards make it easy for new users to quickly add value to the team.

Enterprise connectivity

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

Data profiling

Tightly integrated with DataFlux technology, SAS Data Integration Server includes powerful data profiling capabilities that provide the ability to analyze and assess the quality of data within single or across multiple source systems. With data profiling, you can determine where potential problems exist and what efforts will be required to rectify them. This enables organizations to focus on the root cause of data quality issues, helping them to more accurately plan and better execute their data integration processes.

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 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.

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. 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. 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 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.

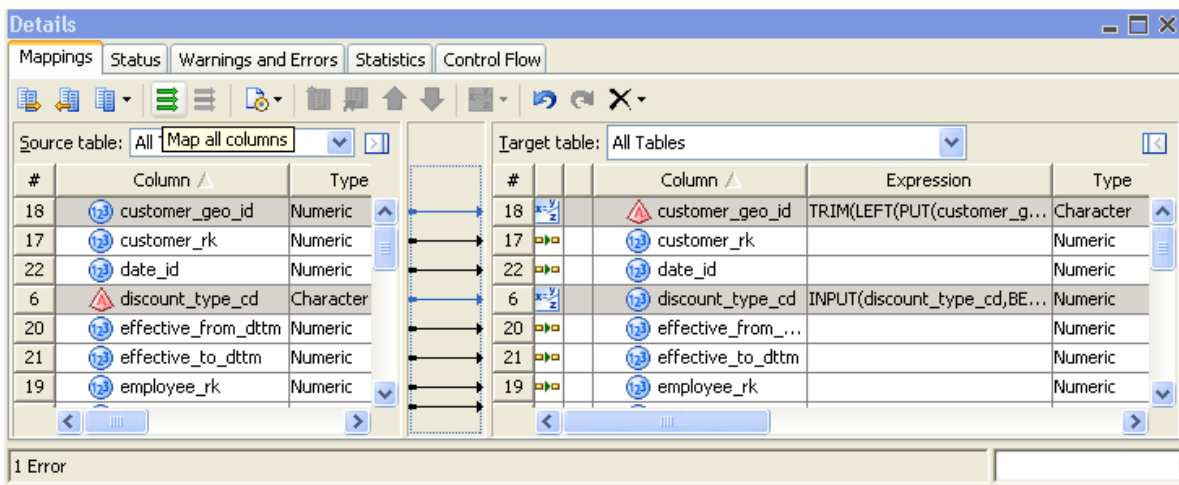
Data federation

Fast access to the most current operational data drives the success of functions with real-time reporting needs, such as call centers and fraud detection activities. In addition, business users need self-service reporting across multiple data sources. SAS 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.

Integrated support services

The features and capabilities provided by SAS Data Integration Server can be modularized into an array of supporting data integration services. These supporting services all leverage 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. These services include scheduling, connectivity, administration, data enrichment, data movement, grid support, data parsing, data cleansing, data auditing, data profiling, data lineage and data transformation.



With SAS 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

- 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 and identify and repair 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 (checkpointed), enabling interrupted processes to be restarted at the point where they were interrupted.
- Powerful DataFlux data profiling.
- 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.

The screenshot displays the SAS Data Integration Server GUI. The top portion shows a data flow diagram with three data objects: 'CUSTOMERS', 'Sort', and 'Append'. Arrows indicate the flow of data from 'CUSTOMERS' to 'Sort', and from 'Sort' to 'Append'. A yellow note box contains the text: 'This job will append incoming data to the customer tables and put the data back into the customer table'. Below the diagram, an 'Errors' dialog box is open, displaying the message: '- No sort by column specified. Select a sort by column.'

The bottom portion of the screenshot shows the 'Details' panel, which displays a source-to-target mapping table. The source table is 'CUSTOMERS (CUSTOMERS)' and the target table is 'Sort (WSNWFE1C)'. The mapping table shows the following columns and their corresponding data types and lengths:

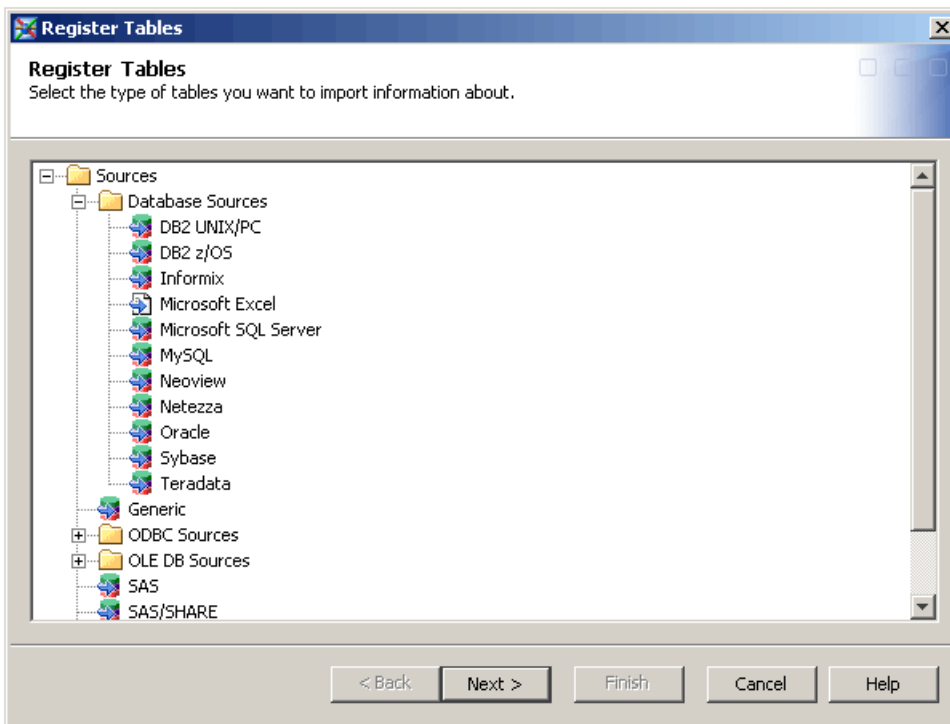
#	Column	Column Description	Type	Length	Inf
1	checking		Numeric	8	(None)
2	duration		Numeric	8	(None)
3	employed	description	Numeric	8	(None)
4	installp		Numeric	8	(None)
5	marital		Numeric	8	(None)
6	coapp		Numeric	8	(None)
7	resident		Numeric	8	(None)
8	property		Numeric	8	(None)
9	age		Numeric	8	(None)
10	other		Numeric	8	(None)
11	existcr		Numeric	8	(None)

(Top portion) SAS 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, join 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.

(Bottom portion) The GUI also shows information such as source-to-target mappings.

Enterprise connectivity

- Database sources: AS400, ODBC, IBM DB2/UDB, Informix, Microsoft Access, Microsoft Excel, Microsoft SQL Server, MySQL, Netezza, DATAlegro, HP NeoView, OLE/DB, Oracle, Sybase, Sybase ASIQ, SAS data sets, SAS® Scalable Performance Data Server®, SAS Scalable Performance Data Engine, Teradata, Greenplum Aster Data and more.
- Mainframe data sources (OS/390 and z/OS): ADABAS, CA-Datcom, CA-IDMS, COBOL, IBM DB2, IMS-DL/I, ISAM files, Oracle, SYSTEM 2000, Teradata, VSAM (KSDS and ESDS), and other file formats.
- Enterprise applications: Oracle Applications, PeopleSoft, SAP BW, SAP R/3, Siebel and Salesforce.com.
- File formats: text, standard and delimited; XML and standard flat files; COBOL Copybooks; and FTP and URL-based. Reads and writes other external data representations such as ASCII, Binary, EBCDIC, Hexadecimal and Octal.
- Direct support for message-oriented middleware, including WebSphere MQ from IBM and MSMQ from Microsoft. In addition, Tibco's Rendezvous is supported with SAS Integration Technologies (included).
- 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.



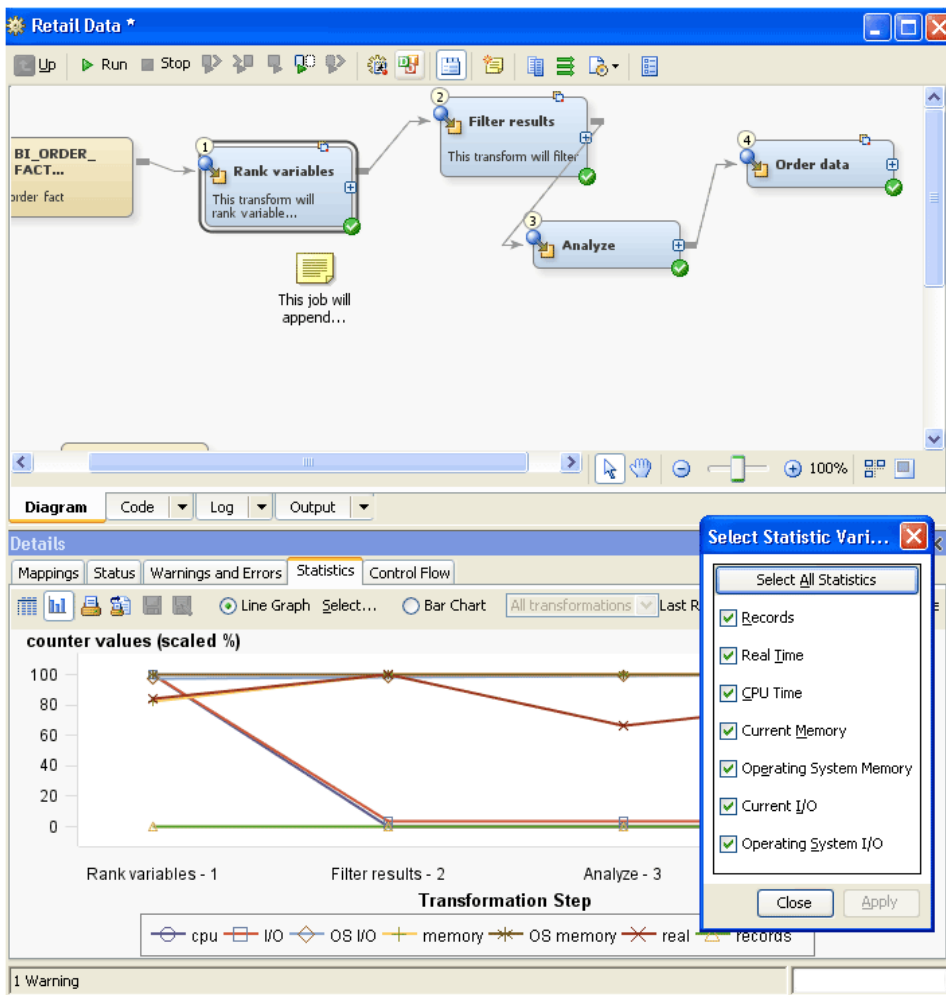
The Register Tables wizard makes it easy to access data from many different systems, as well as read and manage metadata from external sources.

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

- Transformation library with more than 300 predefined table and column-level transformations.
- Transformation Generator wizard or Java plug-in design templates for creating reusable and repeatable transformations that are tracked and registered in metadata.
- Framework for publishing information to archives, a publishing channel, e-mail 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 or as Web services for use by other applications as part of a service-oriented architecture.
- Optimized loaders designed to leverage the various load techniques and bulk-loading utilities.
- A specially designed loader to provide 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 (reverse impact analysis) is critical for both validating processes that are working correctly and for building user confidence in data.
- Multiuser collaboration support includes object check-in and check-out and DEV/TEST/PROD promotion management.
- Wizard-driven metadata importing and exporting.
- Metadata-driven deployment flexibility. Using process metadata, process jobs can be deployed for batch execution, as reusable stored processes and as Web services.
- Metadata reports that provide an easy-to-understand view of the metadata in HTML format for developers, designers and business users to view the content.



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.

Migration and synchronizations

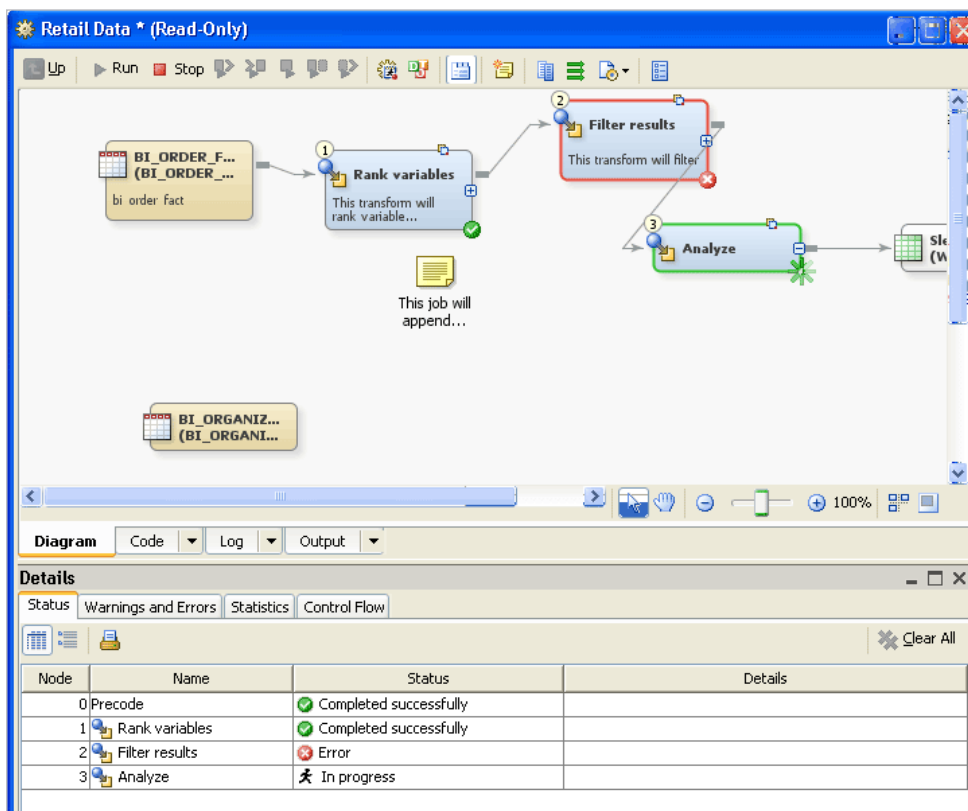
- Metadata-driven access to sources and targets.
- Extensive library of predefined transformations for migration and synchronization that can be extended and shared with other integration processes.
- Embedded, reusable data quality business rules can clean data as it is moved.
- Ability to migrate or synchronize data between database structures, enterprise applications, mainframe legacy files, text, XML, message queues and a host of other sources.
- Integrated scheduler allows changes made in one or more systems to be propagated on a scheduled basis out to other systems in a reliable manner.
- In addition to SAS' own change data capture (CDC) capabilities, when combined with technology from partner Attunity, enhanced CDC recognizes changes to key fields and replicates/synchronizes changes across multiple heterogeneous databases. CDC capabilities provided by Oracle and DB2 are also supported.

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 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.

Integrated support services

- Features and capabilities provided by SAS Data Integration Server can be modularized into 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.
- These supporting services leverage the integrated SAS metadata environment and administrative functions.



Real-time status indicators show specifically which components are actively running, which are finished running and which ones are yet to run, both in the process flow diagram as well as by component in a tabular list. It also shows the associated traffic lighting, which uses real-time visual indicators to show which components of the process have been completed, with and without errors.

SAS® Data Integration Server Technical Requirements

Client environment

SAS Data Integration clients include SAS Data Integration Studio, SAS Management Console and DataFlux® dffPower® Profile. The following client operating systems are supported:

SAS Data Integration Studio

- Microsoft Windows (x86-32):
Windows XP Professional, Windows Vista*, Windows Server 2003 family
- Microsoft Windows on x64 (EM64T/AMD64): Windows XP Professional for x64, Windows Vista* for x64, Windows Server 2003 for x64

SAS Data Quality components

- Microsoft Windows (x86-32):
Windows XP Professional, Windows Vista*, Windows Server 2003 family

SAS Management Console

- AIX: Version 5.3 and Version 6.1 on POWER architectures
- HP-UX PA-RISC: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- HP-UX Itanium: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- Linux for x86 (x86-32): RHEL 4 and 5, SuSE SLES 9 and 10
- Linux for x64 (EM64T/AMD64): RHEL 4 and 5, SuSE SLES 9 and 10
- Microsoft Windows (x86-32):
Windows XP Professional, Windows Vista*, Windows Server 2003 family
- Microsoft Windows (on Itanium):
Windows Server 2003 for Itanium-based systems
- Solaris on SPARC: Version 9, 10

- Solaris on x64: Version 10
- z/OS: V1R7, V1R8, V1R9 and higher
NOTE: Windows Vista Editions that are supported include Enterprise, Business and Ultimate

Supported Web browsers

- Internet Explorer 6 on Windows XP Pro
- Internet Explorer 7 on Windows XP Pro and Windows Vista*
- Firefox 2.0 on Windows XP Pro, Windows Vista* and Linux x86 (SuSE and RHEL)

Server environment

SAS Data Integration Server

- AIX: Version 5.3 and Version 6.1 on POWER architectures
- HP-UX PA-RISC: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- HP-UX Itanium: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- Linux for x86 (x86-32): RHEL 4 and 5, SuSE SLES 9 and 10
- Linux for x64 (EM64T/AMD64): RHEL 4 and 5, SuSE SLES 9 and 10
- Microsoft Windows (x86-32):
Windows XP Professional, Windows Vista*, Windows Server 2003 family
- Microsoft Windows on x64 (EM64T/AMD64): Windows XP Professional for x64, Windows Vista* for x64, Windows Server 2003 for x64
- Microsoft Windows (on Itanium):
Windows Server 2003 for Itanium-based systems
- Solaris on SPARC: Version 9, 10
- Solaris on x64: Version 10
- z/OS: V1R7, V1R8, V1R9 and higher

SAS Metadata Server for SAS Data Integration Server

- AIX: Version 5.3 and Version 6.1 on POWER architectures
- HP-UX PA-RISC: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- HP-UX Itanium: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- Linux for x86 (x86-32): RHEL 4 and 5, SuSE SLES 9 and 10
- Linux for x64 (EM64T/AMD64): RHEL 4 and 5, SuSE SLES 9 and 10
- Microsoft Windows (x86-32):
Windows XP Professional, Windows Vista*, Windows Server 2003 family
- Microsoft Windows on x64 (EM64T/AMD64): Windows XP Professional for x64, Windows Vista* for x64, Windows Server 2003 for x64
- Solaris on SPARC: Version 9, 10
- Solaris on x64: Version 10
- z/OS: V1R7, V1R8, V1R9 and higher
NOTE: Windows Vista Editions that are supported include Enterprise, Business and Ultimate

Optional products or components

- Platform suite for SAS
- Additional SAS Metadata Bridges
- Additional DataFlux data quality components



SAS Institute Inc. World Headquarters +1 919 677 8000

To contact your local SAS office, please visit: www.sas.com/offices

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2009, SAS Institute Inc. All rights reserved. 104166_543125.0809