

SAS® Data Management

Transform raw data into a valuable business asset



To make smart business moves, you not only need sound operations, you need insight. But that's not as easy as it sounds. With more data - and more complex data - arriving every day, how can you make day-to-day decisions and get an accurate picture from your analytics and reporting tools without a foundation of trusted data?

Spreadsheets, emails, reports, customer information, vendor records, operational events and other types of data can too easily become background noise. Add to that the complexity of an IT environment with emerging data sources like Hadoop colliding with traditional data sources, like Oracle or Microsoft SQL Server. Your data challenges of the past just got trickier.

SAS Data Management helps you make sense of this, turning big data into big value. You get data access, data quality, data integration and data governance all from a single platform. Now you can spend less time maintaining your information - and more time running your business.

What does SAS® Data Management do?

SAS Data Management helps transform, integrate, govern and secure data while improving its overall quality and reliability.

Why is SAS® Data Management important?

Whether it's traditional data in operational systems or big data in a Hadoop cluster, data is an asset that every organization has. And managing that data is no longer a convenience - it's a necessity. SAS Data Management is the answer to solving your data integration and data quality challenges.

For whom is SAS® Data Management designed?

SAS Data Management is designed for IT organizations that need to address performance and functional improvements in their data management infrastructure.

Key Benefits

- **Access the data you need**, regardless of where your data is stored - from legacy systems to Hadoop.
- **Improve productivity, reduce development and maintenance time.** Centralized storage, management and reuse keeps data updated and ready for use. Build, document and collaborate on work - and bring new team members up to speed quickly.
- **Govern and secure your data.** Secured authorization and auditing tools add security to your data management program, reducing risk exposure.
- **Work faster and meet time constraints.** Use grid-enabled load balancing and multithreaded parallel processing to rapidly transform and move data, or update with SQL pass-through for popular MPP databases.
- **Eliminate overlapping, redundant tools.** A unified, complete enterprise data management platform eliminates piecemeal management of multiple technologies, helping to control costs.

Product Overview

With SAS Data Management, you can handle a wide variety of data challenges - from efficient processing of big data to accessing and integrating legacy sources - all in a single platform, with in-memory and in-database performance improvements helping to deliver trusted information.

Interactive data integration development environment

SAS Data Management encodes highly technical capabilities in an intuitive, point-and-click graphical user interface. From a role-based console, collaborative efforts are centralized under administrative settings, with wizards for data integration and data management project definition that are controlled by secured permissions and audit traceability.

Integrated process designer

A visual, end-to-end job designer gives full control over coordinated execution processing with advanced specifications for parallel, triggered and conditional jobs based on internal or external events. You can also apply decision management flows to your data management jobs through an integration with SAS Decision Manager.

Connectivity and data access

With SAS Data Management, you can connect to virtually any data, including big data or streaming data, across virtually any hardware environment. You'll benefit from unparalleled, real-time connectivity in batch or through message queues, with specialized table loaders to provide optimized bulk loading for Oracle, Teradata and DB2. Also, load data into SAS® LASR™ Analytic Server for use in SAS Visual Analytics.

Metadata management

Technical, business, process and administrative metadata is stored and managed to facilitate reuse of existing table definitions, business rules and more. Sophisticated mapping technologies make it easy to propagate column definitions from sources to targets and create automated, intelligent table joins across both SAS and third-party data integration and data modeling tools.

Data cleansing and enrichment

SAS Data Management embeds data quality into batch, near-real-time and real-time processes - available through message queues and web services - so

you don't have to question the integrity of your business information. Out-of-the-box standardization rules conform data to your corporate standards, or you can customize rules for specific situations.

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

Simplify collaboration and reuse with more than 300 out-of-the-box SQL-based transforms for creating tables - or to join, insert, delete, update and merge data. SAS Data Management allows you to use analytic transformations for unique insights into data, minimizing data movement and extending transformation power while metadata is documented throughout the entire process.

Migration and synchronization

Whether from mergers, acquisitions or business growth, disparate systems are common and can wreak havoc on your data consistency. SAS Data Management alleviates migration and synchronization issues with embedded, reusable data quality rules and real-time data services for database structures, enterprise applications,

mainframe legacy files, SML, message queues and other sources.

Data federation

Provide consistent business views across all data sources with optimized query processing that provides instant access to information. With SAS Data Management, you can query and use data across multiple systems without physically reconciling or moving source data. It's a quick, cost-effective way to provide access for your business users.

Master data management support

By using features such as semantic data descriptions and sophisticated fuzzy matching, you can check and control data integrity within a web-based reference management interface. You can also identify, standardize and correct master data by each transaction, in hundreds of transactions at a time or in a single pass of the source data.

Data governance

Implement business rules and policies across the enterprise to ensure compliance. These actions can be tracked and monitored across the entire governed environment.

Message queuing

Asynchronous business processes and optimized access for each message queue can reduce the cost of disruptions with minimal administrative effort. With interfaces to the leading message-queuing products as well as event-based application integration, automated action triggers across applications are provided.

Enhanced administration and monitoring

Manage and monitor your complete integration environment, including data integration jobs, data jobs, federation cache jobs, process flows, process jobs and SAS Stored Processes.

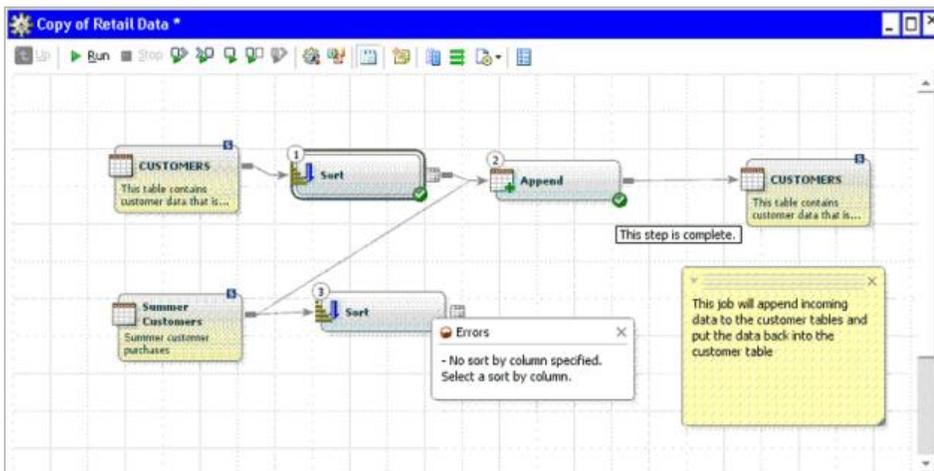


Figure 1. Simplify job building with prebuilt icons into a self-documenting diagram flow. Use the extensive library of out-of-the-box transformations and draw arrows to connect object icons. Yellow sticky notes help clarify job details.

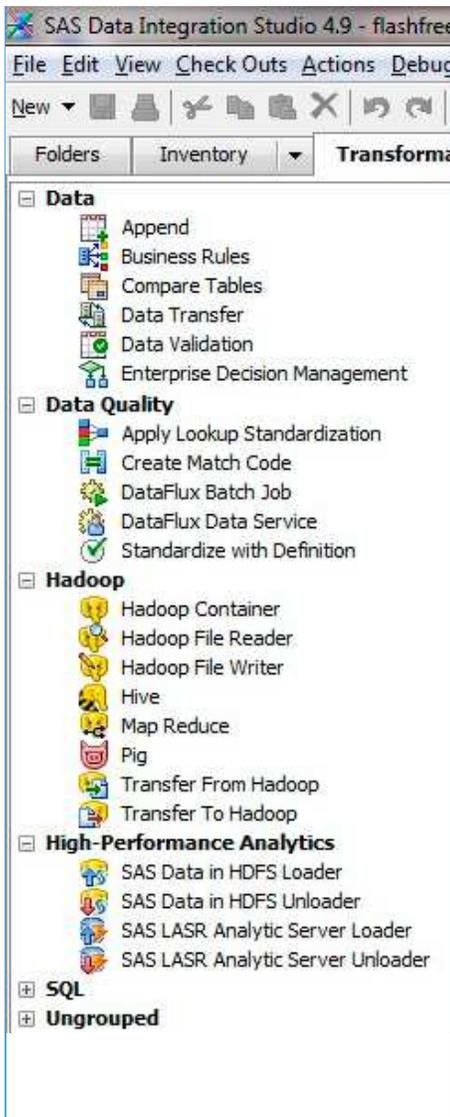


Figure 2. Get access to more than 90 prebuilt data transformations.

Key Features

Interactive data integration development environment

- An easy-to-use, point-and-click, role-based GUI with an intuitive set of configurable windows for managing authorized processes. Drag-and-drop functionality eliminates the need for programming.
- Wizard access to source systems, creating target structures, import and export metadata functions, and build/execute ETL and ELT process flows.
- Customizable metadata tree views let you display, visualize and understand metadata.
- Dedicated GUI for profiling data makes it easy to repair source system issues while retaining the business rules for use in other data management processes.
- Interactive debugging and testing of jobs during development and full access to logs is supported.
- Audit history and check-in/check-out allows designers to see which jobs or tables were changed, when and by whom.
- Ability to distribute data integration tasks across any platform and to virtually connect 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.
- SAS code import capabilities give current SAS users an easy way to import their SAS jobs and code.
- Command-line deployment tool lets users batch deploy many jobs at once using a simple command-line interface.

Integrated process designer

- Build and edit data management processes with a visual, end-to-end event designer.
- Control the execution of data integration, SAS Stored Processes and data quality jobs.
- Conditionally execute jobs based on IF THEN logic and parameterization.
- Fork jobs and processes to execute in parallel.
- Publish job inputs and outputs for parameterized jobs.
- Listen for internal and external events and conditionally raise events.
- Execute external OS-level commands, such as call shell scripts.
- Call REST and SOAP web services.
- List and open old versions of jobs (in read-only mode) and make historical versions current with built-in versioning.
- Provide full support for promotion/migration of jobs in support of DEV/TEST/PROD.
- Use common scripting languages to deploy data integration batch jobs in an automated manner with automated job deployment.
- Execute decision management through a node that supports SAS Decision Manager.

Connectivity and data access

- Allows data transfers to and from SAS® Viya™ via SAS Cloud Analytic Services (CAS). You can continue to use an existing SAS 9.4 platform while beginning to use new features of SAS Viya.
- The S3 transformation lets users push and pull data from an Amazon S3 environment.
- Integration with SAS Data Loader for Hadoop provides a bridge between the ad hoc world of self-service data preparation and governed, operational deployment.
- Provides connectivity in batch or in real time to more data sources on more platforms than most other solutions.
- Data access engines are available for enterprise applications, nonrelational databases, RDBMSs, data warehouse appliances, PC file formats and more.

Key Features (continued)

- Specialized table loaders provide optimized bulk loading of Oracle, Teradata and DB2.
- File reader/writer available for Hadoop file system (HDFS) and support for Hadoop's MapReduce, Pig and Hive within flows as well as Hortonworks.
- Source designer, available for Impala and Pivotal HAWQ, allows SAS Data Management users to see table metadata in Hadoop. The OS/soft PI LIBNAME engine can register PI tables in the source designer and can read and write PI tables.
- A complete and shared metadata environment provides consistent data definition across all data sources.
- Native access methods deliver top performance, reduce data movement 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 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.
- Support for MPP databases - including Aster Data nCluster, Pivotal Greenplum and Sybase IQ - enables more ELT pushdown and support for bulk-load utilities.
- Native support for SQL-based processing.
- Includes connectivity to Aster Data, Pivotal Greenplum, Hadoop and Sybase IQ databases with the ability to push down more processing to the databases.
- SAS LASR Analytic Server transformation options allow use of SASIOLA engine.

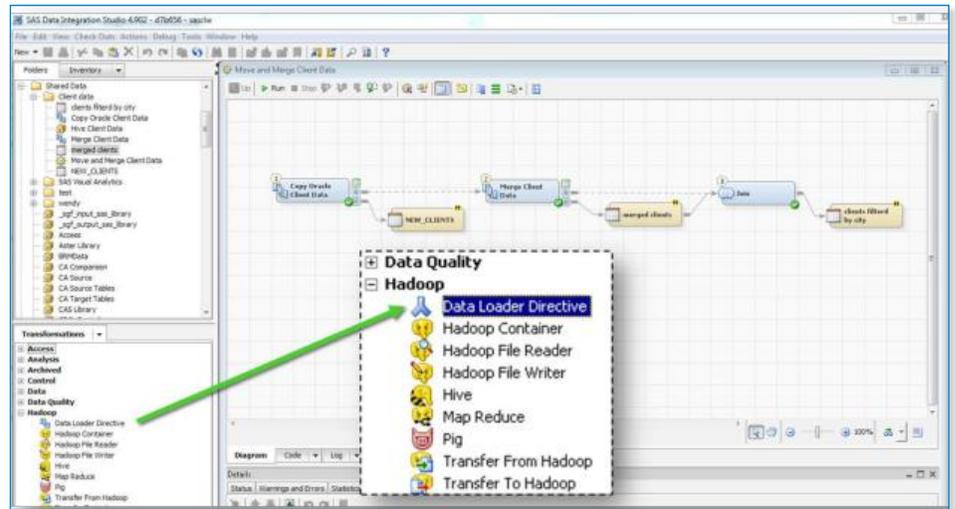


Figure 3. Use Data Loader for Hadoop within an existing SAS environment to blend self-service data prep with the enterprise.

Metadata management

- Metadata is captured and documented throughout transformations and data integration processes, and is available for immediate reuse.
 - Sophisticated metadata mapping technologies quickly propagate column definitions from sources to targets, and create automated, intelligent table joins.
 - Metadata search enables quick location of desired components.
 - 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), critical to validate dependencies, helps build user confidence in data.
 - Batch updates are made to the SAS object metadata repository via the SAS relationship service, for integration with SAS Lineage.
 - 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 as well as column standardization.
 - Metadata-driven deployment flexibility so process jobs can be deployed for batch execution as reusable stored processes or web services.
 - SAS Metadata Bridge enables users to import third-party metadata, including third-party ETL jobs, data transformation logic and data models.
 - SAS Metadata Bridge for Big Data: Apache Hadoop Hive Server, Apache Hadoop Web HCatalog, Apache Hadoop HDFS, Apache Hadoop WebHDFS, Cloudera Enterprise (Hadoop Hive), Cloudera Impala (Hadoop Hive), Hortonworks (Hadoop Hive), MapR Hadoop Hive, Google BigQuery.
 - SAS Metadata Bridge for Tableau: Tableau (file), Tableau Server (repository).

Key Features (continued)

Data cleansing and enrichment

- Data quality is embedded into batch, near-real-time and real-time processes.
- Data cleansing is provided in native languages with specific language awareness and localizations for more than 38 regions worldwide.
- Data quality functions are available in both operational and reporting (transaction and batch) environments.
- An interactive GUI enables you to profile operational data to identify incomplete, inaccurate or ambiguous data.
- Customizable and reusable data quality business rules can be accessed directly within process job flows.
- 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.
- Value can be added to existing data by generating and appending postal addresses, geocoding, 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.

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

- A powerful, easy-to-use transformation user interface that supports collaboration, reuse of processes and common metadata.
- Out-of-the-box, SQL-based transforms deliver ELT capabilities, including create tables, join, insert rows, delete rows, update rows,

merge, SQL set, extract and SQL execute.

- 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.

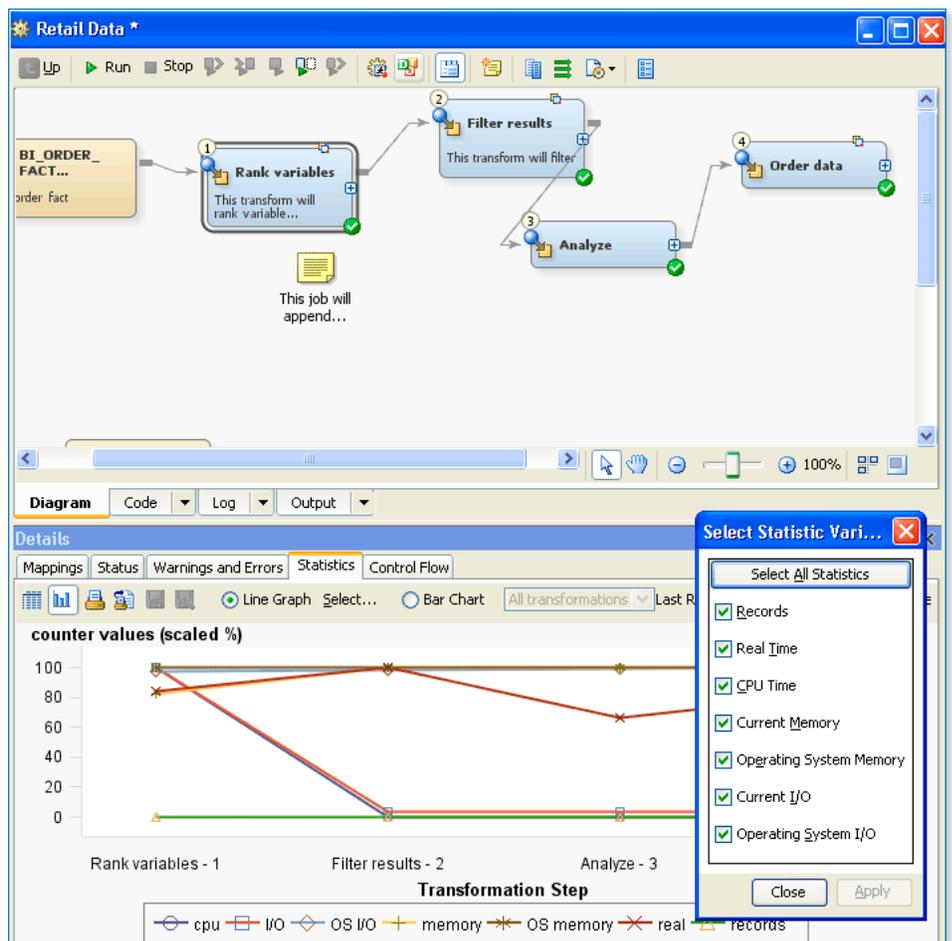


Figure 4. Successful job completion is noted by green check marks for each stage of the job. Metadata statistics are saved for each applied transformation and can be viewed in graphical or tabular (lower right) format to help fine-tune performance.

Key Features (continued)

- Transformation wizard or Java plug-in design templates let you easily generate reusable and repeatable transformations that are tracked and registered in metadata.
- Transformation processes, callable through custom exits, message queues and web services, are reusable in different projects and environments.
- Transformations can be executed interactively and scheduled to run in batch at set times or based on events that trigger execution.
- Framework environment 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.
- Transformations automatically generate high-performance SAS code that is designed for rapid and efficient processing.
- Transformations include: Type 1 SCD support for merge and hash techniques, table differencing and enhancements for Type 2 SCD loaders.
- The Compare Tables transformation compares two data sources and detects changes in data.
- Provides the ability to call REST or SOAP web services.

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.

The screenshot shows the SAS Business Data Network interface. At the top, there is a menu bar with 'File' and 'Help'. Below it is a toolbar with icons for 'New Term...', 'Open', 'Tags...', and 'Lineage'. The main area displays a table with columns: Term Name, Description, Type, Importance, and Stat. The table lists various terms related to educational settings, such as 'ActualEducationalSettings', 'ActualGradeList', and 'Address_Type'. Below the table, there is a detailed view for the selected term 'ActualEducationalSettings', showing its 'Description' and 'History'.

Term Name	Description	Type	Importance	Stat
ActualEducationalSettings	This is a summary of actual educational settings of...	Default	★★★★	Not Spec
ActualGradeList	This is a summary of actual grades instructed by a...	Default	★★★★	Not Spec
ActualGradeListSearchable	This is a summary of actual grades instructed by a...	Default	★★★★	Not Spec
Address_Type	Table that contains address type information from t...	Default	★★★★	Not Spec
Address_TypeCode	Address type code.	Default	★★★★	Not Spec
Address_TypeDesc	Provides a description of the address type.	Default	★★★★	Not Spec
Address_TypeId	Unique identifier for each address type code.	Default	★★★★	Not Spec
Address_TypeName	Name of address type.	Default	★★★★	Not Spec
AgreementNumber	This is the agreement number (formerly referred to...	Default	★★★★	Not Spec
AuthorizedEducationalSettings	This is a summary of authorized educational settin...	Default	★★★★	Not Spec
AuthorizedGradeList	This is a summary of the grades authorized to be l...	Default	★★★★	Not Spec
AuthorizedGradeListSearchable	This is a summary of the grades authorized to be l...	Default	★★★★	Not Spec
AYPDate	Not used.	Default	★★★★	Not Spec

Figure 5. SAS Business Data Network allows business and IT users to collaborate on the creation of business terms, establish links to the terms in a hierarchy of reference data and examine related data elements.

- 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 or 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.

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 improve master data quality.

Key Features (continued)

- Sophisticated fuzzy-matching technology and 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 management 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.

Data governance

- Enhanced, web-based reference data management and business data environment to ease governance and semantic reference, respectively.
- Integrated business glossary allows business terms to be organized hierarchically and related to term owners as well as technical metadata, such as tables and data management processes.
- Publish metadata changes to the SAS relationship service, allowing you to view relationships and evaluate the impact of these changes within the patented SAS Lineage software.
- Extensive data stewardship capabilities, including web-based dashboarding and business rule exception monitoring, for reporting and remediation.

Message queuing

- Integration of asynchronous business processes via message-based connectivity.

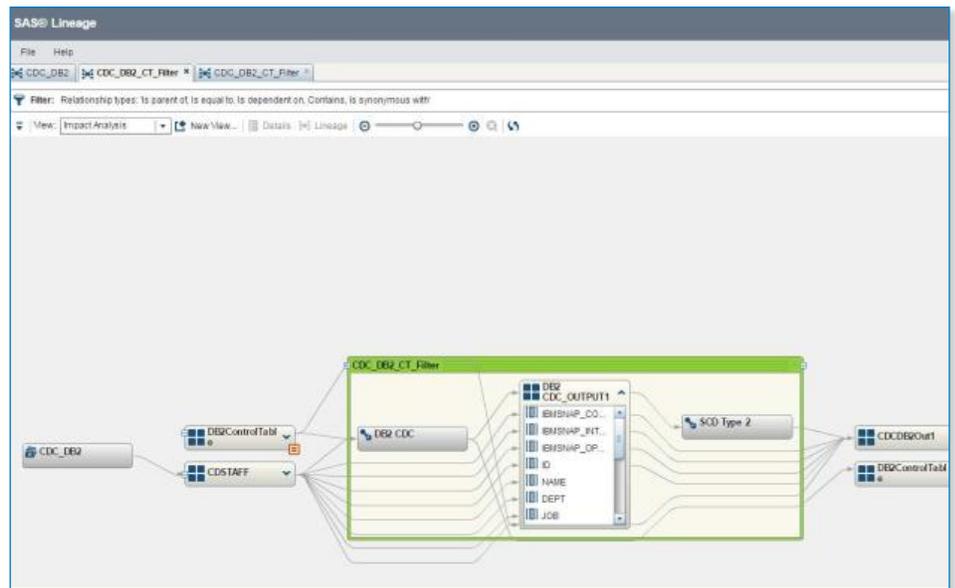


Figure 6. SAS Lineage capabilities help you visualize technical and business metadata from both SAS and third-party sources and tools - and see how a change in one area can affect other data elements.

- 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 run streams and alerts.
- Ability to send and receive messages between distributed and disparate systems.

Enhanced administration and monitoring

- Job status and performance reports and trending information provide the ability to track metrics - such as CPU use, memory and I/O - and deliver updates on how recent process runs perform relative to previous runs.
- Enables users to manage and monitor their complete integration environments, including the following types of jobs and activities:
 - o Data integration jobs.
 - o Data quality jobs.
 - o Federation cache jobs - scheduled queries to update the federation cache.
 - o Process flows.
 - o Access log files from a central, web-based panel for faster, easier troubleshooting.
 - o SAS® Stored Processes.

TO LEARN MORE »

To learn more about SAS Data Management, download white papers, view screenshots and see other related material, please visit sas.com/sasdm.



Pinnacle Solutions, Inc.
152 E. Washington St., Indianapolis, IN 46204
<http://thepinnaclesolutions.com>
Sales 317.423.9143 Ext 1

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

