



SAS® Data Quality Solution

Ensuring information excellence through data quality lifecycle management

What does SAS® Data Quality Solution do?

SAS Data Quality Solution provides an enterprise solution for profiling, cleansing, augmenting and integrating data to create consistent, reliable information that significantly improves returns on corporate intelligence initiatives.

Why is SAS® Data Quality Solution important?

Billions of dollars are lost annually because of poor data quality. Introducing even minor data quality disciplines can provide tremendous savings. SAS Data Quality Solution enables you to improve the accuracy of information your organization receives, reduce operating costs and increase the value of strategic initiatives.

For whom is SAS® Data Quality Solution designed?

SAS Data Quality Solution is designed for two primary groups: business users, including business analysts, data analysts and data stewards, who are responsible for owning the business rules and data relationships as well as analyzing data for business intelligence; and technical users, such as data and database administrators, warehouse managers and other IT professionals, who complete the data cleansing process and ensure that the cleansed data adheres to corporate standards and is consistent across the enterprise.

Inaccurate, inconsistent, invalid and unreliable data affects all organizations, and with data volumes constantly on the rise, it is no wonder that improving data quality has become a key concern.

Many factors affect IT's ability to accurately consolidate data and provide reliable information: increasing numbers of systems and standards, third-party data that is not easily integrated, and duplicate data and applications. In many cases, data standards are not available or are not followed consistently throughout the enterprise.

Many organizations are unaware of their data quality problems. It often takes a catastrophic event for business users to realize the extent of their data quality issues. When strategic corporate projects fail to produce expected returns or fail all together, the problem can often be traced to problems with the data.

SAS provides a powerful enterprise data quality solution for profiling, cleansing, augmenting and integrating data to significantly improve returns on corporate intelligence initiatives.

By integrating data quality within data integration and business intelligence projects, organizations can transform and combine disparate data, remove inaccuracies, standardize on common values, parse values and cleanse dirty data to create a strategic, trustworthy, valuable asset that enhances decision-making power.

SAS Data Quality Solution ensures information excellence through data quality lifecycle management.

Key benefits

- Profile, monitor and actively manage the quality of enterprise data.** SAS Data Quality Solution provides the ability to analyze and assess the quality of data across the enterprise. Profiling allows you to determine where potential problems exist and what efforts will be required to rectify them. Organizations can then focus on root cause analysis, improving key data areas and processes for better planning and project execution.
- Integrate and standardize data across multiple systems and business units.** With SAS Data Quality Solution, organizations can incorporate data quality business rules across data sources and platforms. By implementing standard and custom processes, data from different systems can be regulated and standardized into a unified, accurate view.
- Define data correction rules to reflect organizational changes and cleanse data.** Specialized interfaces make it easy for business analysts and data stewards to create data quality improvement processes and visualize the impact of business rules and data cleansing efforts. State-of-the-art data quality tools enable business and technical users to cleanse, standardize, integrate and augment data, and can be customized to meet the requirements of each organization.
- Provide decision makers with trustworthy information.** Accurate, consistent, valid and reliable data delivers consistent reporting and analytic results. By automatically integrating data quality into data integration processes, data used for business intelligence and analytics will be current and accurate.



Product overview

SAS Data Quality Solution provides a complete set of data quality tools and easy-to-use interfaces designed to meet the needs of both business and technical users. The tools include the ability to profile data to uncover data discrepancies and determine the effort required to rectify them, an easy-to-use interface for defining business rules, and a platform-independent server environment to execute the rules on data from any platform in any format.

When poor data quality is identified, the solution provides the ability to cleanse and augment that data to ensure consistency and accuracy. By providing the ability to customize the algorithms used to parse and cleanse data based on language-specific constructs or individual user requirements, SAS Data Quality Solution addresses the needs of each customer and extends the value of any strategic solution.

Powerful, easy-to-use interface

Business users and data stewards can work in a Windows-friendly environment to analyze their data, define business rules and create data quality process improvement specifications for IT users. This enables them to visualize the impact of poor data and easily define processes that are repeatable and reusable to improve data quality.

Data profiling

Data profiling gives organizations a better understanding of data quality issues that exist within data structure, content and relationships and enables better planning and project execution. SAS Data Quality Solution provides the ability to profile and assess the quality of data across the enterprise. A robust environment analyzes data across the enterprise to determine nuances and discrepancies. An easy-to-use interface and an interactive reporting mechanism make it easy to determine areas of poor data quality and the amount of effort required to rectify them.

Cleansing and standardization

Easy-to-use tools enable data stewards, business users and technical users to analyze and prototype data quality cleansing processes and apply corrections to improve the accuracy of analysis. You can parse data values (i.e., name parts, address parts, e-mail addresses and any free-form text values), apply address standardizations and validate address data based on local standards.

Matching and deduplication

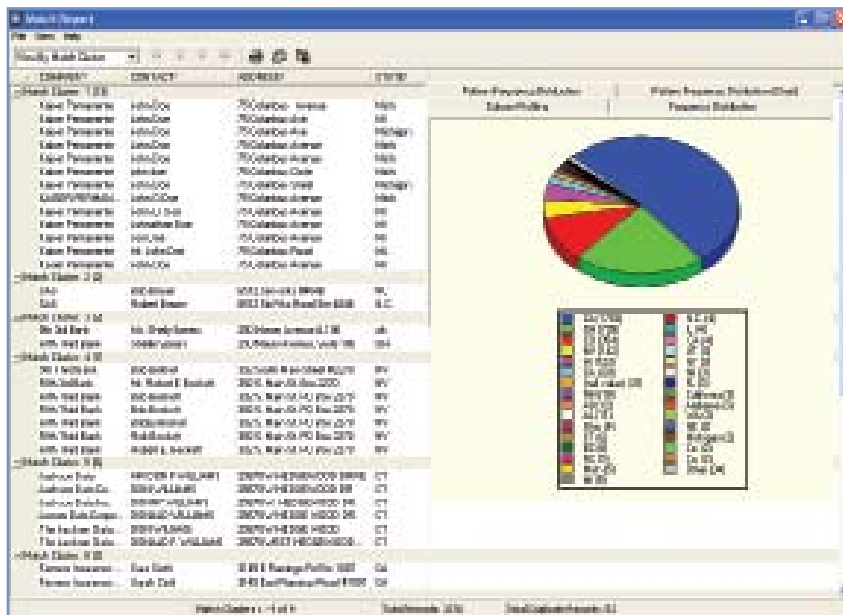
Matching algorithms can join dissimilar data from multiple sources using algorithms that include heuristics and multinational data phonetics. This helps eliminate guesswork when complete matches are not possible and creates a consistent view of information. Unique key values are created with fuzzy logic to group together information with similar values (for example, Robert, Bob and Bobby) across one field or multiple fields. You can remove and merge duplicate values in data to significantly reduce storage requirements and provide consistent information across data sources.

Identification analysis

This capability determines the gender of an individual, which may be helpful in segmenting data for targeted marketing purposes. It can also determine whether a value is a person or an organization, which could be used to determine the type of services to offer when a call is placed to customer service. The algorithms have been extended to also identify significant pieces of contact information (name, address, city, state, ZIP code, Social Security number, account number, etc.).

Customization

Personalizing or customizing the parsing, matching, standardization and identification algorithms and rules



SAS Data Quality Solution provides an easy-to-use interface that lets you uncover inconsistencies and inaccuracies in your data, and its robust server environment gives you the power to analyze data quality across your entire organization.

provides the ability to control the data quality process based on an individual organization's business requirements. For example, rules can be enhanced or created to control how product codes, quantities and other characteristics are parsed from a string of data. A common Quality Knowledge Base lets you share this information as well as leverage language-specific algorithms between server and client components.

International support

SAS Data Quality Solution understands the differences in languages, including names, addresses and other business data. For example, an address in Germany does not contain the same data as an American address and is not parsed the same way. SAS Data Quality Solution provides the ability to correctly standardize data in multiple languages. It also supports Unicode and Double Byte Character Set data.

A component of SAS® Data Integration

SAS Data Quality Solution can be applied to any data source supported by SAS on any supported platform, which ensures enterprise scalability and enhances the quality of all data integration processes and strategies. The collaborative approach to data integration and data quality helps eliminate duplication of effort by data stewards and analysts, and their work can be further leveraged by IT professionals to enforce the quality metrics over time and across the enterprise.

Extends the value of SAS® Business Intelligence solutions

Integrating data quality lifecycle management capabilities within business intelligence solutions provides high-impact results and ensures an acceptable return on investment.

Key Features

Powerful, easy-to-use graphical user interface

- Windows-friendly environment lets users analyze data, define business rules and create data standardization and data matching/integration specifications for IT users.

Data profiling capabilities

- A robust environment analyzes enterprise data to determine nuances, discrepancies and inaccuracies in data.
- Provides an easy-to-use interface to determine areas of poor data quality and the amount of effort required to rectify them.

Data cleansing, deduplication and standardization

- Provides the ability to eliminate or reduce inconsistencies in corporate data.
- Allows duplicate records to be removed from a customer, supplier or product database or from other corporate information.

Match code generation (single or composite element keys)

- Consolidates data from multiple data sources when a complete match is not possible.
- Creates unique key values by using fuzzy logic to group together information with similar values (for example, Robert, Bob, Bobby, etc.) across one field or multiple fields.

Householding and site identification

- Allows the creation of householding or site ID based on one or more conditions.

Customization of parsing, standardization and matching algorithms

- Personalize parsing, matching and standardization algorithms.
- Create or enhance the rules used to parse the parts of the names, addresses, e-mail addresses, product codes and other business data.
- Add rules to matching algorithms about which part of a data string weighs the heaviest in a match.
- A Common Quality Knowledge Base (QKB) lets you share this information between SAS products (server) and DataFlux® products (clients).

Identification analysis

- Determines the gender of an individual for targeted marketing purposes.
- Determines whether a value is a person or an organization.

International support with many language- and country-specific constructs

- Understands the nuances in various languages, including names, addresses, organizations and other areas where data parsing is required.
- Employs local knowledge of these data types.

Leverages SAS® software's powerful data integration capabilities

- By integrating with SAS capabilities, SAS Data Quality Solution can be part of an overall data integration strategy.
- Data quality can be applied to all SAS supported data sources and platforms, which ensures enterprise scalability.
- Provides an interactive approach to data quality that eliminates duplication of effort when creating business rules.

Extends the value of SAS® business intelligence and analytics

- Integrating data quality capabilities within business intelligence solutions creates high-impact results, ensuring an acceptable return on investment.

Address verification (CASS and SERP certification)

- Takes advantage of postal service discounts (i.e., US, Canada, etc.) for mass mailings with certified addresses. (Add-on functionality.)

SAS® Data Quality Solution **Technical Requirements**

Client environment

- Windows (x86-32): Windows XP Professional, Windows Vista*

Server environment

- 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, SuSE SLES 9
- Linux for x64 (EM64T/AMD64): RHEL 4, SuSE SLES 9
- Microsoft Windows (x86-32): 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.

Required software

- Base SAS®

Optional software

- SAS® Data Integration Server (for data quality transformation plug-ins)
- SAS/ACCESS® software (for access to third-party databases)
- Additional Quality Knowledge Base (QKB) locales
- dfPower Verify (CASS, SERP, US Geocode, Canadian Geocode, phone and ZIP code analysis, and world address verification)
- dfPower Monitor (real-time business rule monitoring)
- dfPower Explorer (metadata profiling)

SAS Data Quality Solution is also included as part of SAS Enterprise Data Integration Server.



**THE
POWER
TO KNOW.**

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 © 2008, SAS Institute Inc. All rights reserved. 101422_502021.0708