SAS® Data Quality for Midsize Business
Get accurate, reliable and consistent data

What does SAS® Data Quality for Midsize Business do?
SAS Data Quality for Midsize Business helps companies assess, improve, monitor and manage the quality of all their data – structured and unstructured – so they can rely on it to make strategic decisions and improve core operational processes. Designed and priced for small and midsize businesses, the easy-to-use solution meets the needs of both business and IT users. It is available in desktop and client/server environments so you can start small and grow as your needs change.

Why is SAS® Data Quality for Midsize Business important?
The solution provides an easy-to-use, unified interface for all data quality and data integration initiatives that helps midsize businesses more effectively use their disparate data sources while maintaining data quality. It includes a visual, point-and-click interface that lets users run data quality initiatives without any manual coding. The solution promotes collaboration between IT and business users, helping them agree upon and implement common data definitions and business rules for improved data consistency and reliability.

For whom is SAS® Data Quality for Midsize Business intended?
It is designed for business and IT users at small and midsize organizations that need easy-to-use, affordable and quick-to-deploy solutions to support comprehensive data quality initiatives.

Small and midsize organizations are under pressure from all sides. Today’s difficult economic environment is marked by extreme competition, and includes larger players who are fighting to keep their market share. Small and midsize businesses can no longer rely on intuition or past successes to propel their businesses forward. To foster business success, they must use their organizational data effectively.

Unfortunately, data is spread across multiple systems. Many organizations struggle to get a comprehensive, accurate view of all this data. When strategic projects fail to produce expected returns or compliance efforts are compromised by reporting inaccuracies, the issue can often be traced to problems with the data. With SAS Data Quality for Midsize Business, both business and IT users can:

• Assess and improve the quality and reliability of organizational data by profiling, cleansing, augmenting and standardizing it.
• Establish uniform data quality standards and controls to ensure regulatory or legislative compliance and enhance data governance efforts.
• Get started with a master data management (MDM) initiative without needing to acquire and integrate additional technologies.

Key Benefits
• Get a complete, accurate and unified view of data. The solution connects disparate technology systems and streamlines information transfers to give users a precise, comprehensive view of all data – structured and unstructured. An automated early-warning system flags out-of-tolerance data.
• Make faster, accurate decisions. Through a common, shared interface, IT and business users collaborate to define centrally maintained business rules, data definitions and data standards. IT oversees and validates data, keeping information current, consistent and reliable.
• Free IT to focus on strategic business activities. With a centralized platform and uniform standards, IT is freed from repetitive maintenance tasks. You can save and reuse data integration and data quality jobs and workflows, and both business and IT can use the point-and-click interface to visually explore data for quality issues and create workflows.
• Mitigate risks and reduce costs. Industry-leading data profiling, data quality and entity resolution technologies help users assess the scope and nature of data quality problems across multiple systems and sources. Consistent standards and business rules reduce or eliminate costs of duplicate or incorrect data storage and administration. A consolidated portfolio of capabilities from one vendor minimizes licensing, integration, maintenance, training and support costs.
• Choose a solution that grows with you. The flexible, extensible solution is available in both desktop and server environments so you can start small and grow as your needs change.
Solution Overview

SAS Data Quality for Midsize Business is a comprehensive solution that gives your organization a complete set of capabilities for data quality assessment, improvement and document extraction. Use it to integrate your structured data with unstructured data from emails, old records, invoices, claim forms and more – for a full, reliable view of your customers, products and other entities. The solution includes the foundation for accelerated MDM deployment so you can get started on an MDM initiative without acquiring any additional technologies.

Designed and priced for small and midsize businesses, the solution’s easy-to-use, interactive interface lets users create data quality and integration workflows in a point-and-click environment – no manual coding required. To enhance productivity and outcomes, users can visually explore data for anomalies. Using a shared interface, IT and business users can collaboratively define business rules, data definitions and standards. This minimizes duplicated efforts and helps improve data consistency and reliability.

Data profiling

Data profiling is critical when you need to identify the root cause of poor quality among disparate data sources. SAS Data Quality for Midsize Business includes industry-leading data profiling capabilities to provide a more granular assessment of your data’s strengths and weaknesses. You can connect to virtually any data source, including flat files, relational databases and mainframe systems; extract the data from the source system; and perform an in-depth assessment of your organization’s data, examining the structure, completeness and suitability of data assets.

Through a data discovery program, the solution helps you gain the knowledge to design effective data quality, data integration and master data management business rules to support your organization’s data-driven initiatives.

Data monitoring

The solution includes data monitoring capabilities to help users detect and present anomalies within the data. You can create business rules once and use them repeatedly across applications, for a uniform set of business standards across all your systems. This approach improves business unit and IT collaboration and helps you rapidly take a data governance program from concept to reality.
Data exploration

Metadata analysis uncovers existing trends and characteristics of your data, helping provide a clear picture of the data types and sources. This understanding is an essential first step to improving data quality. The solution enables you to collect, catalog and organize metadata from virtually any data source anywhere in your organization, from a single interface. You can connect disparate data sources, extract metadata — such as field names, data type and length — and use advanced matching algorithms to identify relationships and potential redundancies.

Data quality

Better data leads to better decisions. SAS Data Quality for Midsize Business includes patented matching technology, transformation routines and identification logic to help you easily correct duplicate records and identify nonstandard data representations and unknown data types. Out-of-the-box data standardization rules eliminate inconsistencies. You can plan and prioritize data correction initiatives, establish data hierarchies and reference data definitions, and validate data for overall improved accuracy.

The solution is fully extensible, allowing the addition of new data definitions. Using a customizable interface, you can teach the solution to parse, match and standardize any type of data, including product names, descriptions, numeric information and much more.

Entity resolution

The ability to link and consolidate entity information with a high level of confidence is critical to data management initiatives. Information about the same customer, product or employee may exist in multiple databases, in many unique forms. The challenge is to find and resolve similar records in different data sources. Entity resolution measures the degree of similarity between

Key Features

Data profiling

- Business rule validation lets you authenticate data against standard statistical measures as well as customized business rules.
- Key relationship analysis helps you discover hidden relationships in data across tables and databases as well as different source systems. Automatically identify orphaned records within a specific database or across databases.
- Redundant data analysis identifies unnecessary information across tables or data sources. The engine supports both exact match and fuzzy match redundancy.
- Data validation verifies that the data in your tables matches appropriate descriptions.
- Statistical analysis establishes trends and commonalities in corporate information and examines numerical trends via mean, median, mode and standard deviation.
- Data sampling — of full data or sample data from a source system — lets you analyze for data quality. Users can specify the sample interval.
- Data profiling results can be exported to HTML, Microsoft Excel or text files, and users can write custom reports.

Data monitoring

- Monitor data to detect anomalies such as variances, values inside or outside of ranges, values that violate mathematical calculations, those that vary from historical values and more.
- Design and enforce rules to determine if data is maintained and within proper control limits, and ensure data meets predefined business rules.
- Create data alerts and controls to verify that data remains in compliance with internal and external policies.
- React to data problems quickly before inaccurate or invalid data has a negative effect.
- Create customized business rules to validate and audit operational processes.

Data exploration

- Connect to disparate enterprise data sources, including relational database management systems, XML files, spreadsheets and text files.
- Extract metadata such as field name, data type and length.
- Collect, catalog and organize metadata.
- Use advanced matching algorithms to identify relationships and potential redundancies, and to compare field names, types and formats.
- Identify and catalog relationships between databases’ tables.

Data quality

- Unify data from different databases and source systems. Merge multiple files or duplicate records from a single file into one entity or record with industry-leading matching technology and innovative clustering logic.
- Establish data hierarchies and reference data definitions to create a unified view of a particular data entity, such as customer or product.
- Perform gender analysis using an advanced gender analysis algorithm.
- Intelligently break multivalue fields into parsed elements such as address, city, first and last name, date, phone number, etc.
- Standardize data and eliminate duplications and inconsistencies by using advanced standardization (mapping) routines such as element, phrase and pattern standardization.
- Replace an original field with a new value or append a standardized value directly onto the source record.
- Support householding through sophisticated match keys that enable the engine to group records and assign an integer-based unique identifier.
- Use natural language parsing to separate values as required.

continued on next page
two data elements, often based on weighted records matching. By defining these unique instances, information can be assigned to a single, consolidated record, or can be flagged for manual intervention or further processing.

Traditionally, entity resolution has been a difficult task because there is rarely an exact correlation between all information in every data field. SAS Data Quality for Midsize Business includes advanced fuzzy-matching technology and identification logic that reconciles data and accurately integrates multiple records.

**Data integration**

Data integration involves combining processes and technologies so your organization can make the most effective use of disparate data sources. This solution provides a unified development and delivery environment that gives you a single interface for data quality and data integration activities – including design, navigation and data management, data integration jobs and workflows. Its document extraction capability gives users a single, complete picture of the data that’s being used for analysis.

**Master data management foundation**

The solution’s easy-to-use interface and integrated approach to data management provide the foundation for accelerated MDM deployment. Users in any industry can get started by taking a phased approach, without needing to acquire and integrate additional technologies.

<table>
<thead>
<tr>
<th>Key Features (continued)</th>
</tr>
</thead>
</table>

**Entity resolution**

- Identify individuals across multiple data sources even from incomplete and unobvious relationships.
- Manage entity resolution routines through advanced fuzzy-matching technology.
- Create multiple-record clusters, confidence scores and scatter plots to determine potential clusters.
- Recognize when slight variations suggest a connection between records.

**Data integration**

- Data access capabilities enable users to connect disparate technology systems and streamline the transfer of business information to and from various technology resources.
- ETL and ELT methods let you extract, transform and load data from multiple sources into a data warehouse using both traditional batch processing and in-database methods.
- Document extraction techniques let users automatically access and integrate unstructured data tucked away in emails, old records and invoices, eliminating the need for manual processing along with associated delays and errors – so complete data can be processed quickly for analysis.
- Data migration means you can transfer data to new or different locations while improving the data’s accuracy and consistency during the migration project.
- Data linking and matching allow you to match information within or across data sources, standardizing formatting differences.

**Master data management foundation**

- Integrate the creation and management of master data resources with comprehensive data management practices.
- Create a hub of master data based on a subset of your existing data, using a phased approach.
- Combine MDM capabilities with matching, clustering and other data management initiatives.
- Conduct batch processing with an architecture that supports many MDM implementations without unnecessary complexity.
- Connect to MDM hubs as if they were any other data target.

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

To learn more about SAS Data Quality for Midsize Business system requirements, download white papers, view screenshots and see other related material, please visit [sas.com/dq-smb](http://sas.com/dq-smb).