Challenge
Organizations must maintain data quality to ensure the accuracy and reliability of operational and transactional processes, as well as business analytics and business intelligence reporting.

Solution
Data quality solutions provide functionality that enables businesses to effectively manage data quality by assessing, transforming, and monitoring data.

Gain Valuable ROI with Data Quality
Quality data is crucial to decision making and planning. Within an organization, acceptable data quality is vital to operational and transactional processes and to the reliability of business analytics and business intelligence reporting. Data quality can be affected by the way data is entered, stored, and managed. Building a proper cloud-based warehouse for the data requires an integrated, single source of data that can then be used to make business decisions. Since the data is usually sourced from several different systems, it is important to make sure that the data is standardized and cleaned before loading into a data warehouse. Some of the benefits of using data quality tools:

- Automatically generates the mappings that businesses can use to correct data. These mappings are based on rules that businesses choose to apply to the data.
- Enabling businesses to include data quality and data profiling as an integral part of the data integration process.
- Storing metadata regarding the quality of the data alongside the actual data. The scope of your metadata management defines how you capture and represent your data lineage—giving you an audit trail which can be as granular or high-level as you like.

Data Profiling and Cleansing
Data profiling is the first step for any business to improve information quality and provide better decisions. Using a robust data analysis method will allow businesses to discover and measure defects in the data before businesses start working with the data. This will enable organizations to automatically correct any inconsistencies, redundancies, and inaccuracies in both the data and metadata. Data profiling also enables organizations to discover many important things about their data. Below are some common elements found when organizations perform data profiling:

- A domain of valid product codes
- Columns that hold the pattern of an e-mail address
- A one-to-many relationship between columns
- Anomalies and outliers within columns
- Relations between tables even if they are not documented in the database

After a data set passes profiling requirements, the next step is validating that it aligns with your business and schema rules for each source record. This ensures that you are processing data to the highest quality standards. The right data cleansing solution offer flexible and efficient techniques to identify and clear or correct both subtle and complex issues that can hide deep within large data sets.
Common data issues that can be caught and corrected through data cleansing include:

- Missing data values
- Data duplicates
- Inconsistent data formats

**Data Auditing**

Data auditing allows you to oversee both current and historical data integration, profiling and cleansing operations in your system. It provides evidence that your data integration processes are in compliance with business, technical and regulatory standards.

Through data auditing you can:

- Track all data integration operations such as updates and deletions.
- Perform data evaluations between sources and destinations.
- Track the success, failure and execution duration of every component in your data integration.

Data auditing is one of the most important steps in ensuring data quality, as it provides insight into the state of your data at all levels of operations.

**Data Quality: How Good is Good Enough?**

Data quality is important because without high-quality data, organizations cannot understand or stay in contact with their customers. It’s easier than ever before to find out key information about current and potential customers. Having this information properly stored can enable businesses to market more effectively, and encourage customer loyalty that can last for a long time.

The highlight of data quality is how well the data supports the context in which it’s consumed. Your legal department, for example, may use “XYZ-Company” while your finance department uses “XZY”, and both records are of equal quality. Quality is a never-ending judgment, one that needs to be defined by the business that’s consuming the data. An essential element of data governance, trustworthy data serves critical business needs across the enterprise from the legal department to finance to marketing and literally everywhere else within the organization. Driving data quality requires a repeatable process that should include:

- Defining the specific requirements for “good data,” wherever it’s used.
- Establishing rules for certifying the quality of that data.
- Integrating those rules into an existing workflow to both test and allow for exception handling.
- Continuing to monitor and measure data quality during its lifecycle (usually done by data stewards).

Because rules and needs change and new systems can be added, truly successful data quality initiatives need to scale to address those new requirements.

**Looking for The Right Data Quality Solution**

When looking for a data quality solution, it’s recommend that organizations put the following at the top of their “must-haves” list:

**Reliability:** Your business should have a solution with a successful track record as well as a comprehensive approach to ensure that a data quality process supports the key steps around discovering, defining, applying, monitoring, and measuring progress.
Agility: In addition to dealing with greater volumes of data, organizations must be able to adapt to new sources and types of data. With a flexible and scalable data quality system, you can apply agile methods to your data projects and solutions to: increase efficiency and shorten time to data business use, build custom solutions to meet your business goals, simplify data quality development processes, and easily adapt processes to align with new business requirements and objectives.

Portability: Whether your organization uses legacy mainframes or the latest technology, you’ll want a data quality tool that can evolve as your business does. That means the data quality solution you deploy should allow you to scale across platforms, from on-premises to the cloud.

Flexibility: Your data quality solution should be flexible enough to meet the requirements of your business, rather than forcing the business to adapt itself to rigid technology constraints.

Benefits of Data Quality

To get a better understanding of why data quality is important, you only need to look at the many benefits that accurate and actionable data gives to organizations.

Improved Customer Relations: Accurate data will likely improve relations with your customers. This enables organizations to know their customers which helps organizations to anticipate and meet their needs. This creates a great deal of goodwill with the customer base, another reason why data quality is important.

More Consistent Data: Larger companies and organizations that offer several points of entry for their customers must continually face the problem of data that is inconsistent across the business. Inconsistent data leads to duplicate mailings, it keeps company and organizational departments from reaching key clients. Data quality important because it helps keep every department in the organization on the same page when it comes to analyzing and meeting the needs of their customers.

More Effective Marketing: The importance of data quality is evident in marketing initiatives. In the past, the lack of demographic and other important data about customers meant that companies could only market to the broadest audience. The richness of demographic information available today enables marketing that is more focused and more likely to achieve marketing goals.

Lower Mailing Costs: Accurate customer data reduces the amount of undeliverable mail, which saves organizations money in postage costs in having to resend packages that would have arrived the first time. Additionally, companies that make the most concerted efforts to work with the U.S. Postal Service to keep their customer addresses up to date will receive discounted postage rates for direct mail and shipping.

Conclusion

A key to data quality in an age of self-service analytics is to make IT a partner to the business. Data quality is not about locking everything down to a few users, but about enabling broad groups of users with the appropriate controls. Business users get secure access to shared production data sources managed by IT, so they don’t have to go searching for the right data. IT can set security controls and business practices, helping users keep their data accurate and secure. When done right, data quality is also about providing the right data to the right people whenever and wherever they need it.

Visit aws.amazon.com/mp/mdm to learn more about Business Intelligence on AWS.

Get Started with Business Intelligence Software at AWS Marketplace

- Find and deploy the solution you need in minutes
- Save money with pay-as-you-go pricing
- Scale globally across all AWS regions

Visit aws.amazon.com/mp/mdm