

Data Monitoring: Add Controls to Your Data Governance Program



Many companies today have come to realize the substantial impact that its underlying data can have on every aspect of a company's life. From corporate strategy to customer service to supply chain management, the information that you gather on customers, prospects, products, inventory or employees can be a major factor in your organization's ability to properly understand its current corporate landscape.

In fact, good data holds the key to better decision-making, as it gives you the background to assess your business situation and locate trends that can give you a competitive edge. Good data helps you respond quickly to shifts in customer demands or to changes in the supply chain. And due to the rise of data governance – the overall management of the accuracy and security of corporate information – the quest for good data is now a vital component to any corporate IT initiative.

But an organization cannot simply cleanse and improve the quality of its data, then expect this data to be a static resource over time. Data is a fluid, dynamic and ever-changing resource, so data quality is not a once-and-done activity. The need for data quality never goes away. It requires continual oversight.

Data, in essence, reflects the changing world around you. Customer records become obsolete as people move or switch jobs. Catalogs for products and supplies become outdated. Without a commitment to ongoing data quality, the integrity of an organization's data quickly becomes incorrect or invalid as it reaches core applications.

Data monitoring has become a key component of a complete data quality integration practice, giving you the tools you need to understand how and when your data strays from its intended purpose. Monitoring also helps you identify and correct these inefficiencies through the automated, ongoing enforcement of customizable business rules. Data monitoring ensures that once your data becomes consistent, accurate and reliable, it remains that way over time, giving you confidence when making information-based decisions for your organization.

The Role of Data Monitoring

Building better information resources is an arduous task for any organization. Improvement typically happens gradually and requires difficult and time-consuming processes to understand specific data quality issues – and correct these issues.

To start any data improvement initiative, you have to identify and fix the problems that created bad data if you are to reap the rewards of high-quality data. For most organizations, these objectives require a cross-section of resources from multiple departments, often taking months, if not years, to correct.

Unfortunately, many organizations believe that once they have improved existing data sources, their data quality problems are cured. And indeed, by improving data quality in a one-time project, they have improved the quality of their decision-making. However, as companies accumulate new data from a wide variety of sources over time, the overall quality of data most likely will suffer. Quality data takes constant vigilance.

Data monitoring may include simple data profiling trend analysis over time, or it may include specific analysis of complex business rules. By configuring rules that define acceptable data quality values, monitoring can be used to automatically identify records that violate quality standards and alert users of the violations.

*Quality
data
takes
constant
vigilance.*

Monitoring supports ongoing data governance efforts to ensure that all incoming and existing data meets pre-set business rules. If you have set up a data improvement or data governance program already, you are ahead of the game, and your organization is headed in the right direction. You are already able to recognize data problems, inspect data sources and data processes, and implement the data and process corrections needed to get the problems fixed.

By adding a data monitoring component to that program, you greatly enhance your data quality efforts and make your data much more reliable. By continually checking and re-checking the integrity and viability of your corporate data, you can make sure that high-quality data remains consistent, accurate and reliable. Data monitoring builds on your initial data quality initiatives by providing the technology needed to make sure your data continues to provide the most value for your organization.

Set Controls on Your Data

Traditionally, data quality meant implementing a narrow set of data cleansing and matching procedures such as finding and eliminating duplicate records or correcting customer names and addresses. Data quality was a tactical effort, typically employed within a department or business unit – and it rarely, if ever, received executive support or acknowledgement.

Today, the need for validating and correcting this data – and the detailed rules that oversee this process – is more important than ever. The amount of new data pouring in from multiple, disparate data sources is growing exponentially as organizations, along with their data, merge and consolidate in the new global economy.

With these and other factors directly affecting your own information, you need to understand if the data that you have in your applications and databases is still doing what you expect it to do. Is your data still valid? Is it still meeting its intended use? For almost all businesses today, the answer to both of these questions is no. And given the growing number of compliance initiatives you now face – from internal data governance programs to government-driven regulations – you need greater control over your data, now more than ever.

A data monitoring solution can help you maintain consistent, accurate and reliable data by helping you:

- Identify trends in data quality metrics and data values
- Provide instant alerts for violations of pre-established business rules
- Quantify the costs associated with data quality and business rule violations
- Detect variances from cyclical runs
- Recognize when data exceeds pre-set limits, allowing you to immediately update this data by addressing the problems upfront before the quality of your data declines

Data monitoring gives you the assurance to know that once you fix your data problems, they will remain within tight limits. When data does get out of control, you know immediately – and can address problems before the quality of your data declines. Data monitoring allows you to confidently reclaim control of your company's data, not let it control you.

Data monitoring asks a simple question. Is your data doing what you expect it to do?

Enforcing Business Rules for Data Monitoring

The business reasons for implementing a data quality program are changing. Once, companies used data quality to cleanse customer data that could help drive sales and marketing efforts. Although, this is a worthy goal, companies have begun to ask their data quality tools to perform more in-depth examinations of data. The result is the adoption of business rules monitoring capabilities, which allows users to establish and enforce complex, configurable standards for data.

By monitoring business rules, you can:

- Enforce data governance standards
- Audit and validate the quality of your company's data on an ongoing basis
- Create customized rules to streamline your operational processes
- Identify and fix operational inefficiencies with a continual, automated enforcement of pre-set rules
- Improve the accuracy of your business processes
- Understand and refine those processes by logging exceptions and violations to business rules in a repository, allowing you to see and address trends

Both IT and business users can create and enforce customized rules to monitor any data-driven, enterprise wide initiative. These business rules include detecting unexpected variances, validating numeric calculations, testing domain values against a list of acceptable values, or validating data before allowing a transaction to continue.

Once you discover violations, the software automatically notifies other designated users of rule violations and corrects the data. The monitoring engine also copies the violated data to a repository for further examination and correction. The ability to create a number of different outcomes (write to a file, create an alert, start a corrective routine, etc.) is extremely powerful, as it gives you the flexibility to create complex business rules to manage virtually any situation.

With these capabilities, the business rules engine becomes a tool that ensures that data meets organizational standards. In doing so, these rules become a critical control mechanism within data governance or operational compliance efforts.

Figure 1 shows a business rules manager for a data monitoring console. This depicts how rules are created and organized within a data quality environment. The screen not only shows the name of the rule, but also details the fields impacted by the rule – as well as any other tasks that utilize the rule.

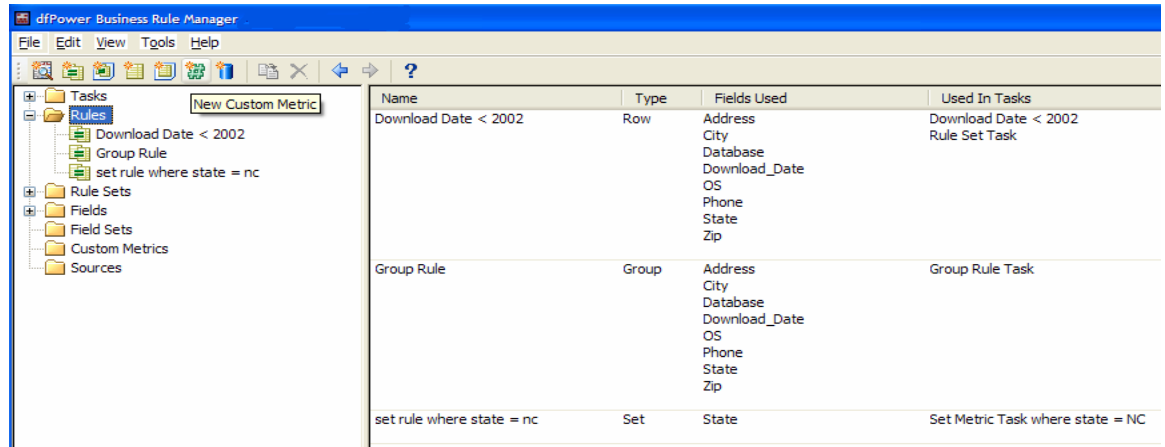


Figure 1 - A business rules manager allows any user to create and document rules for data quality.

A business rule manager also allows companies to reuse business rules in both batch and real-time environments. You can apply rules once or utilize them to validate information as it flows throughout the enterprise.

While the business rules manager provides a creation and editing platform, the process requires a repository (see Figure 2) to store and manage all created business rules. This structure allows users to track and trend specific rules and their violations. Business users can view rule exceptions and track violations over time. And, through a service oriented architecture (SOA), users can expose these rules as Web services along with traditional data quality services.

The screenshot shows the 'dfPower Monitor Viewer' application. It displays a table with the following columns: Date, Task*, Rule, # Triggers, # Rows Processed, User ID, and Data Source. The table contains the following data:

Date	Task*	Rule	# Triggers	# Rows Processed	User ID	Data Source
Age Test						
02/14/06 04:40:14	Age Test	Age test	1	5000	none	none
02/14/06 04:38:10	Age Test	Age test	1	5000	none	none
Base Rider Check						
02/14/06 05:17:16	Base Rider Check	Base Rider Calc	1	5000	none	none
02/14/06 05:16:09	Base Rider Check	Base Rider Calc	0	5000	none	none
Policy Number Check						
02/14/06 11:32:55	Policy Number Check	Policy Number Check	1	5000	none	none
02/14/06 11:31:58	Policy Number Check	Policy Number Check	1	5000	none	none
02/14/06 11:30:56	Policy Number Check	Policy Number Check	0	5000	none	none
Transaction Type Check						
02/14/06 05:04:04	Transaction Type Check	Transaction Type Check (list)	1	5000	none	none

Figure 2 - A data monitoring repository tracks and trends the application of a specific rule as well as the rule violations.

For any company attempting to gain more control over enterprise data, data monitoring provides the ability to report the changing nature of corporate information over time. Figure 3 demonstrates a dashboard screen, showing the number of triggers or violations experienced by certain business rules.

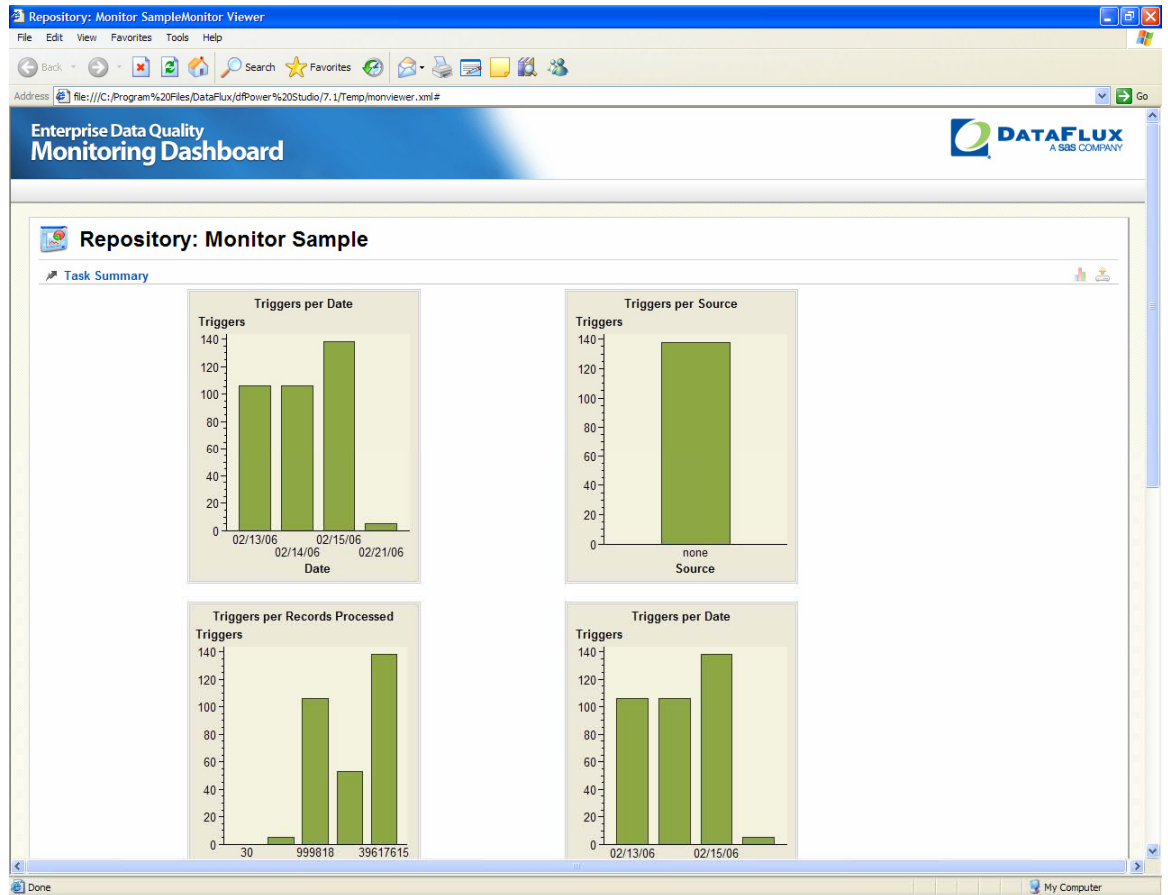


Figure 3 - A dashboard provides high-level information on trends and attributes of the data.

This dashboard provides the type of information that a manager or an executive needs to understand the state of data within the enterprise. These web-based reports provide a baseline of knowledge about any problematic data.

The business analyst or data steward working directly on data quality initiatives would have similar, but more granular screens that provide the understanding required to manage enterprisewide data quality initiatives. These dashboards would fuel much of the work done by these employees, providing an automatic way of managing data quality over time.

Data Monitoring and Process Improvement

With data monitoring, you have the final phase of any data improvement or data governance cycle. By this time, you have profiled the data, created correction and standardization routines, integrated information from multiple sources and enriched the data as necessary. Data monitoring provides you with a critical opportunity to control the quality of this data, thus adding an important piece of any data management program.

However, organizations also understand that the outcome of monitoring routines (reports, alerts, etc.) can provide important insight into the processes that create bad data. By studying this information, you can do more than correct the data; you can also begin to improve the overall efficiency of the enterprise.

The concept of process improvement has been an integrated part of the manufacturing cycle for years. Manufacturers understand that more precise, refined processes create a better (and more profitable) product.

Years ago, Walter Shewhart introduced a process improvement method designed to drive continuous improvement. Initially developed for manufacturing, it has broad business applicability that can help companies increase the quality of any asset – whether a finished product or a set of data. The method, known alternately as PDCA (Plan-Do-Check-Act) or as the Shewhart Cycle, provides an overall framework for improving the precision and overall effectiveness of internal procedures.

The Six Sigma methodology uses a variation of this method, called DMAIC (Define-Measure-Analyze-Improve-Control). PDCA was built primarily for manufacturing improvement; Six Sigma was developed to specifically address customer experience problems, whether the customers existed inside or outside of a manufacturing environment. Other quality improvement processes, such as Total Quality Management (TQM), follow similar themes.

Process improvement methodologies, such as Six Sigma, apply to data quality.

Regardless of the terminology employed by these methodologies, the processes are basically identical in purpose, if not in practice. Most process improvement methodologies can be broken into three major phases:

- Identify the problem
- Fix the problem
- Monitor the problem

Based on these three fundamental process steps, you can implement a comprehensive process for dealing with complex data issues. This process couples the one-time correction of a data source with the process for finding the root cause of the problem. And with data monitoring, you can ensure that once a problem is found and fixed, it stays fixed.

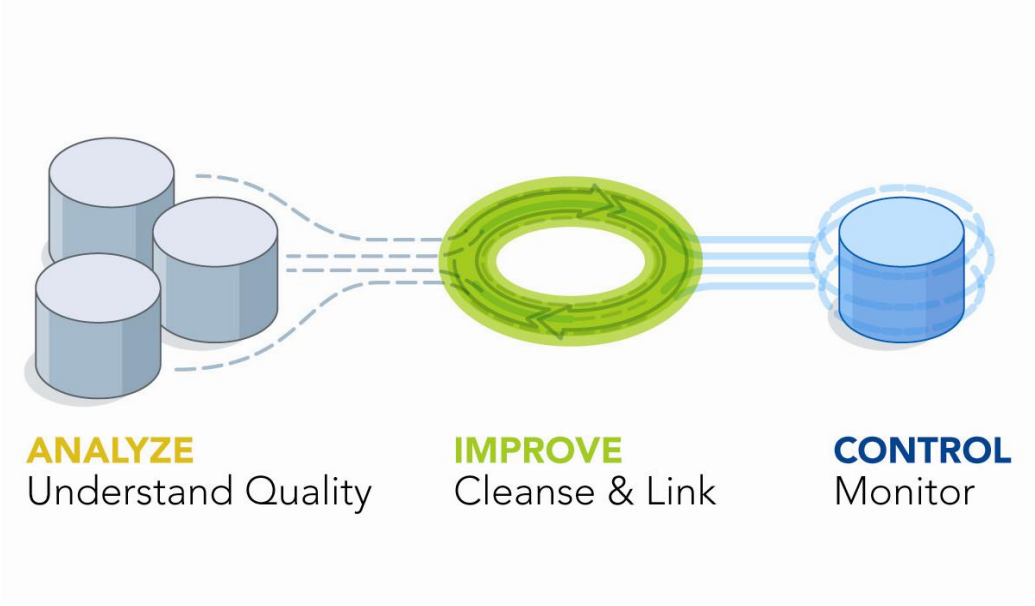


Figure 4: The process flow for data improvement

This methodology is an easy-to-follow process for business users and data analysts as they work to uncover data quality issues in their organization. To solve a problem, you must eliminate the root causes of the problem and not just the problematic result. The Analyze-Improve-Control process contains the essence of PDCA or Six Sigma methodologies and is applicable for any company looking to improve the quality of their data assets.

Analyze

The first step in the data improvement process involves discovering the root cause of the problem and defining a path to improvement. Some problems are the result of a bad business decision, such as a decision that cannot be substantiated with underlying data. Other problems may continuously erode your organization's effectiveness but never surface as an identifiable problem.

Almost all data consolidation or integration projects involve data profiling – also known as data discovery or data inspection – to identify any quality issues that may exist. This initial analysis measures all aspects of data quality, including completeness, accuracy, consistency and duplication of your data. A data profiling effort also measures business rule integrity, as well as the deviation from corporate standards.

Understand strengths and weaknesses of data.

For instance, suppose you have a database with product information, and to ensure customer service, your customer relationship management (CRM) system should list products in the same nomenclature as the product database. Profiling would let you know when the CRM system contains products or product names that do not exist in the product database – and flag the problem for correction.

All of this important information can allow the business user to build a solid roadmap for solving the identified data quality problems. Typically, once the data problems have been identified, you need to conduct a root-cause analysis of the problem. In the example above, you might survey customer relations or sales staff to understand the origination of product data in the CRM system. From that information, you can understand the source of this bad data and correct it in the future.

Improve

Once you identify specific data quality issues, the next phase involves planning and executing processes for the improvement of the data. Here, a business analyst will set up data quality rules to fix all of the identified data anomalies. An improvement phase that typically takes weeks or months can be drastically reduced to days – or even hours – with easily configurable technology and automatic data quality routines.

Create correction routines to address and resolve quality problems.

In the improvement phase, you must correct the existing data sources while also correcting the underlying process that produced the problems. Essentially, you treat the symptom *and* enact improvements to solve the root problem. There are many techniques used for fixing and controlling an error-prone process, including:

- Assuring that edits and validation occur during original data capture
- Assigning clear procedures to improve data entry processes
- Conducting intensive training and performance measure refinement
- Performing data checks as information flows from application to application

Control

After data correction, the next step is to establish a control mechanism to ensure that you can maintain high levels of data quality on an ongoing basis. The control phase builds on the first two phases. From the analysis phase, you know which data issues were problematic or “out of limits” in the past. And the improvement step provides an in-depth look on how to fix those problems. With the control phase, you merge your understanding of the data with the methods to improve data to conduct ongoing checks of data quality.

Enforce business rules to maintain high levels of data quality.

To effectively monitor and control your data, there are a few important processes you need to investigate:

- Ongoing reporting and analysis of potential problem areas
- Alerting mechanism that recognizes out-of-control data records – and automatically flags the data owner or responsible party
- Trending of data to recognize cyclical variations to allow business analysts to view a historical view of data problems

Data monitoring mechanisms can include both high-level reports and detailed, drill-down analysis. For instance, a data monitoring technology might provide a high-level report that shows 20% of the annual revenues in Database A do not match the combined value of the 12 monthly payments in Database B. If you choose, you can drill down into that report to understand which records are out of limits – and see if there is a quick solution to this problem.

Conclusion

More and more, organizations are acknowledging the significant impact that data has on every aspect of a company’s lifecycle and, ultimately, on the bottom line. These organizations have embraced data quality as the means to improve the quality and credibility of their information, in order to make better, more informed business decisions. But true data quality is not a once-and-done initiative. It takes a continual effort to oversee fluid, ever-changing data.

Data monitoring is a vital component of any complete data quality and data integration solutions, giving you the tools you need to inspect, validate and control your company data. Data monitoring helps you make sure that all incoming data meets your standards and automatically alerts your data stewards or IT staff when data violates standards or exceeds limits.

With data monitoring, you can now embrace techniques that will help you maintain the highest data quality standards over the long haul. Business rules monitoring gives you the ability to identify and correct operational inefficiencies by creating, customizing and enforcing rules that allow you to audit and validate the quality of your data on an ongoing basis.