

The Butterfly Effect on Data Quality

How small data quality issues can lead to big consequences

White Paper

Table of Contents

- How a Small Data Error Becomes a Big Problem 3**
- The Pervasiveness of Data 4
- Customer Relationship Management (CRM) 5**
- Benefits..... 6
- ERP / Supply Chain Data 6**
- Benefits..... 7
- Orders / Billing System..... 7**
- Benefits..... 8
- Data Warehouse 8**
- Benefits..... 8
- Essential Functionality..... 9**
- Corporate Cultural Benefits..... 10**
- Talend Data Quality 10**
- Data Profiling 10
- Reporting..... 11
- Data Quality Portal 11
- Features 11
- Cleansing and Standardization 12
- Unified Data Management Platform 12**

“Butterfly Effect”—the idea, used in chaos theory, that a very small difference in the initial state of a physical system can make a significant difference to the state at some later time.

Dictionary.com's 21st Century Lexicon

The term “butterfly effect” refers to the way a minor event - like the movement of a butterfly’s wing - can have a major impact on a complex system - like the weather. The movement of the butterfly wing represents a small change in the initial condition of the system, but it starts a chain of events: moving pollen through the air, which causes a gazelle to sneeze, which triggers a stampede of gazelles, which raises a cloud of dust, which partially blocks the sun, which alters the atmospheric temperature, which ultimately alters the path of a tornado on the other side of the world.

Enterprise data is equally susceptible to the butterfly effect. When poor quality data enters the complex system of enterprise data, even a small error - the transposed letters in a street address or part number - can lead to revenue loss, process inefficiency and failure to comply with industry and government regulations. Organizations depend on the movement and sharing of data throughout the organization, so the impact of data quality errors are costly and far reaching. Data issues often begin with a tiny mistake in one part of the organization, but the butterfly effect can produce disastrous results.

This white paper explores the butterfly effect of poor data quality on CRM, ERP, billing, data warehouse, and other enterprise systems. Although there are many types of data quality problems, this paper uses a simple transposed character - to show how the impact of the error is amplified as it travels throughout the data ecosystem.

How a Small Data Error Becomes a Big Problem

When data moves from place to place within a corporation, it affects systems in different ways. The more interconnected systems you have, the more potential for escalating data quality issues. Without some kind of data management strategy that includes data quality, your company is likely to suffer from multiple ongoing data problems.

Something as simple as a transposed character has lasting effect. For example, if two call center workers enter the same customer address as “25 Main St.” and “25 Mian St.”, it could potentially impact:

- your ability to market effectively to that customer
- your ability to comply with “Do Not Call” regulations by flagging only one of the records
- accurate record keeping on the customer’s transactions
- mailing costs
- customer satisfaction
- revenue and customer reporting activities
- much more

At the highest level, data quality issues affect:

- **Revenue:** Without accurate data on customers, your organization can’t achieve revenue goals. Poor data quality most often affects your ability to reach customers, cross-sell, up-sell and meet their needs/desires.
- **Efficiency:** Untrustworthy data results in wasted time and resources, as you are forced to check and recheck facts and figures before you make decisions. It also prevents data from being easily shared by others in your organization.
- **Compliance:** Poor data affects your ability to comply with industry and local laws, such as Sarbanes-Oxley, Basel II, Do Not Call, and HIPAA. Lack of compliance can lead to unnecessary fines and levies.

The Pervasiveness of Data

When data enters the corporate ecosystem, it rarely stays in one place. Consider the typo in the customer address as it travels throughout your enterprise. Marketing accesses the data in the CRM system to reach customers. A successful campaign results in orders which have an

impact on Shipping, Billing, Supply Chain, Support, and other systems. Finally, a manager will want to aggregate all campaigns and orders to a data warehouse for reporting.

If the data enters the ecosystem as incomplete, incorrect or duplicate, many systems are affected. The possible impact is shown in Figure 1.



Figure 1: Data is pervasive. As it moves throughout a corporation, data impacts systems and business processes. The negative impact of poor data quality reverberates as it crosses departments, business units and cross-functional systems.

Customer Relationship Management (CRM)

A customer database or CRM system typically focuses on name and address data. Inevitably letters will be transposed, names will be comma reversed, and postal codes will be missing, among other errors. Data quality software makes it easy to standardize name, address, and any other contact information needed for a strong relationship with your customer.

In most countries, postal authorities specify how this type of information should be formatted and what it should contain. If you're somewhat precise about following these rules, most mail makes it to its

destination. If you're very precise, you may qualify for postal service discounts.

If you know what a broken item looks like and you know what a fixed item is supposed to look like, you can design and develop processes; train workers and automate solutions to solve real business problems. Knowledge is built into the system, so you don't have to start from scratch every time you need to cleanse CRM data.

Benefits

By standardizing customer data, you will be able to offer better, more personalized customer service. And you will be better able to contact your customers and prospects for cross-sell, up-sell, notification and services.

ERP / Supply Chain Data

An ERP or supply chain system focuses on parts, descriptions and inventory data. Since government agencies don't control the way supply chain and ERP data is defined, you may not have an idea about how the data should look in an ideal state, but it should provide an accurate depiction of the physical warehouse or just-in-time supply chain system. You need to know what is in stock, when it can be supplied and how much it costs.

Holding just the right amount of inventory is crucial to optimizing costs. After all, inventory costs are incurred every hour of every day in areas including warehouse storage, heat and electricity, staffing, product decay and obsolescence. With this in mind, consider the impact of a transposed letter in an ERP system. Someone enters part number XL-56YJM instead of LX-56YJM. Until the error is discovered and corrected, you may hold duplicate parts in inventory or not be aware of parts carrying the slightly different SKU because of the transposed letter.

You also want to take advantage of volume discounts. If the data is unstructured, making it difficult to understand global buy patterns, the company may miss out when negotiating with the vendor.

Because there is no standard format for ERP data, few of the steps for fixing the data are done for you ahead of time. It is critical to establish a methodology for data profiling in order to understand issues and challenges. Since there are few governing bodies for ERP and supply chain data, the corporation and its partners must often come up with an agreed-upon standard, with input from users of the data to understand context, how it's used, and the most desired representation.

Benefits

If you have clean data in your supply chain, you can achieve some tangible benefits. First, the company will have a clear picture about delivery times on orders because of a completely transparent supply chain. Next, you will avoid unnecessary warehouse costs by holding the right amount of inventory in stock. Finally, you will be able to see all the buying patterns and use that information when negotiating supply contracts.

Orders / Billing System

Orders and billing systems leverage multiple data sources to predict build and delivery time, provide correct invoices, and ensure that accurate invoices are sent to customers. When data quality issues arise in this area, you see invoice errors leading to payment delays. The billing system is usually the system of record for credit decisions. Data quality issues can lead to issuing credit to unworthy customers while denying credit to those who deserve it.

Billing systems can also be a primary source for financial information. Accurate data in the billing system can help you quickly finalize

monthly, quarterly or annual reports. If you find yourself checking and rechecking the data, or there is general lack of trust among your employees about the billing data, it could be that data quality problems have crept in.

Benefits

If you have clean data in your billing systems, you can achieve the tangible benefits of more accurate financial reporting and correct invoices that reach the customer in a timely manner. An accurate bill not only leads to trust among workers in the billing department, but customer attrition rates will be lower if invoices are delivered accurately and on time.

Data Warehouse

A data warehouse leverages data from across the organization, ranging from sales information to salaries, operational data, inventory data, etc. By its very nature, data coming into a data warehouse will not share a common data definition, metadata model or common way of representing information on customers, sales, inventory or operations. Identifying necessary transformation and accurately standardizing data is vital.

Decision support systems rely on information in the data warehouse to detect trends and inform strategic decision making. When data quality problems are detected in the data warehouse, it calls into question the facts that decisions are based on.

Benefits

If you have standardized the data feeding into your data warehouse, you can dramatically improve business intelligence. Employees can access the data warehouse and be assured that the data they use for reports, analysis and decision making is accurate. Using the clean data in a

warehouse can help you find trends, see relationships between data, and understand the competition in a new light.

Essential Functionality

An effective data quality solution will include:

- **Data Profiling** - A solid understanding of data issues before you implement a major CRM, ERP, Billing or Data Warehouse project can prevent surprises - like late stage data quality issues just before you're ready to go-live. For example, data profiling could help you see that materials, parts and supplies are designated in more than one way in the database and to check for compliance with a metadata model when upgrading or moving between CRM systems. Data profiling can also check the completeness and integrity of name, address, e-mail, web URL and other client data.
- **Reference Data** - Reference data is used to standardize parts and descriptions as a "look-up" table. Many industries have standard ways to designate parts and descriptions. Data quality tools should help you leverage these standards to apply conformity to your data.
- **Data Standardization** - Data standardization enables you to standardize such things as city names, postal codes, and street names. This makes it easier to reach customers and find duplicates.
- **Matching Technology** - Matching goes a step further than standardization, helping you find duplicates and recognize households and other relationships in your customer records.
- **Monitoring** - Data quality issues are rarely one-and-done. Regular monitoring of key data quality metrics ensures that your reports are

accurate and that your marketing materials and invoices make it to their destination in a timely manner

Corporate Cultural Benefits

How much credibility does your data have? When you lose faith in the integrity of your data, it's hard to generate enthusiasm for ideas and campaigns based on the data, because they seem (and often are) doomed to failure. Conversely, when you trust the data behind your CRM, supply chain and decision support systems, it motivates you to harness the information in new ways, giving rise to fresh ideas and inspiration. Data quality is not just about saving money - though there is a tangible payback when you eliminate inaccuracy and duplication for your information systems. It's about creating new opportunities by harmonizing the data from disparate systems and providing technologies, executives, and business teams with quality data.

Talend Data Quality

With Talend Data Quality, organizations have a single, powerful solution for reporting and monitoring the quality of their data across the enterprise, allowing them to easily align the data quality facets of a data governance program. This alignment improves processes and communication on your valuable data assets. With Talend Data Quality data quality scorecards can be distributed to the data source owners to track data quality improvements over time and drive compliance with enterprise-wide data quality standards.

Data Profiling

Data profiling provides both an evaluation of your current state of data quality and measurement of it over time. You can generate informative

reports and share key data quality metrics with your team - the first step in solving data quality issues.

Reporting

Talend Data Quality includes report generation where data stewards can leverage the data profiling results to create predefined reports that watch for violation of data quality thresholds. The reports can be generated in Adobe Acrobat and other formats and published via e-mail, wiki or intranet site.

Data Quality Portal

Talend Data Quality delivers customized, key quality indicators to a web-based portal where teams can collaborate on the process of improving data quality across the enterprise. Data Quality Portal provides an easy-to-use, browser-based view into key data quality metrics. The portal opens up the issue of data quality to a broader audience in your organization, fostering positive change around the way your company manages data.

Features

- Customizable web-based dashboards that display critical information on data quality processes needed by business users
- Personalized, default dashboards for different levels within the organization that include alerts, data views, and links
- Access control so users and managers only see the dashboards they need for their job
- Alerts that update in real time so you're always on top of data quality issues
- Track progress against realistic targets on an attribute-by-attribute basis
- Export reports in popular formats like PDF, XLS, XML and HTML

Cleansing and Standardization

Talend Data Quality also has powerful tools for repairing and cleansing data. Talend Data Quality allows you to use internal or external reference data to set the standards for values, including regular expressions to set standards for data shape and size.

Set up cleansing processes using a wide range of dedicated data integration and quality components. Nicknames or addresses are for example easily normalized and standardized using dedicated components. These dedicated components, such as name and address cleansing, fuzzy deduplication, third-party address validation and standardization services, are natively available in Talend Data Quality.

Unified Data Management Platform

Talend is an open source data management company offering complete, end-to-end data management tools. Data quality provides the foundation for building a complete data management strategy, but data management is not limited to data quality. Data management requires that you be able to access the data regardless of the source system, manage metadata and provide the transformations required for data integration. Data management is also about process change, the kinds of changes you make as you move toward master data management and a more mature data governance strategy. Talend tools provide an easy path as your data management needs mature.