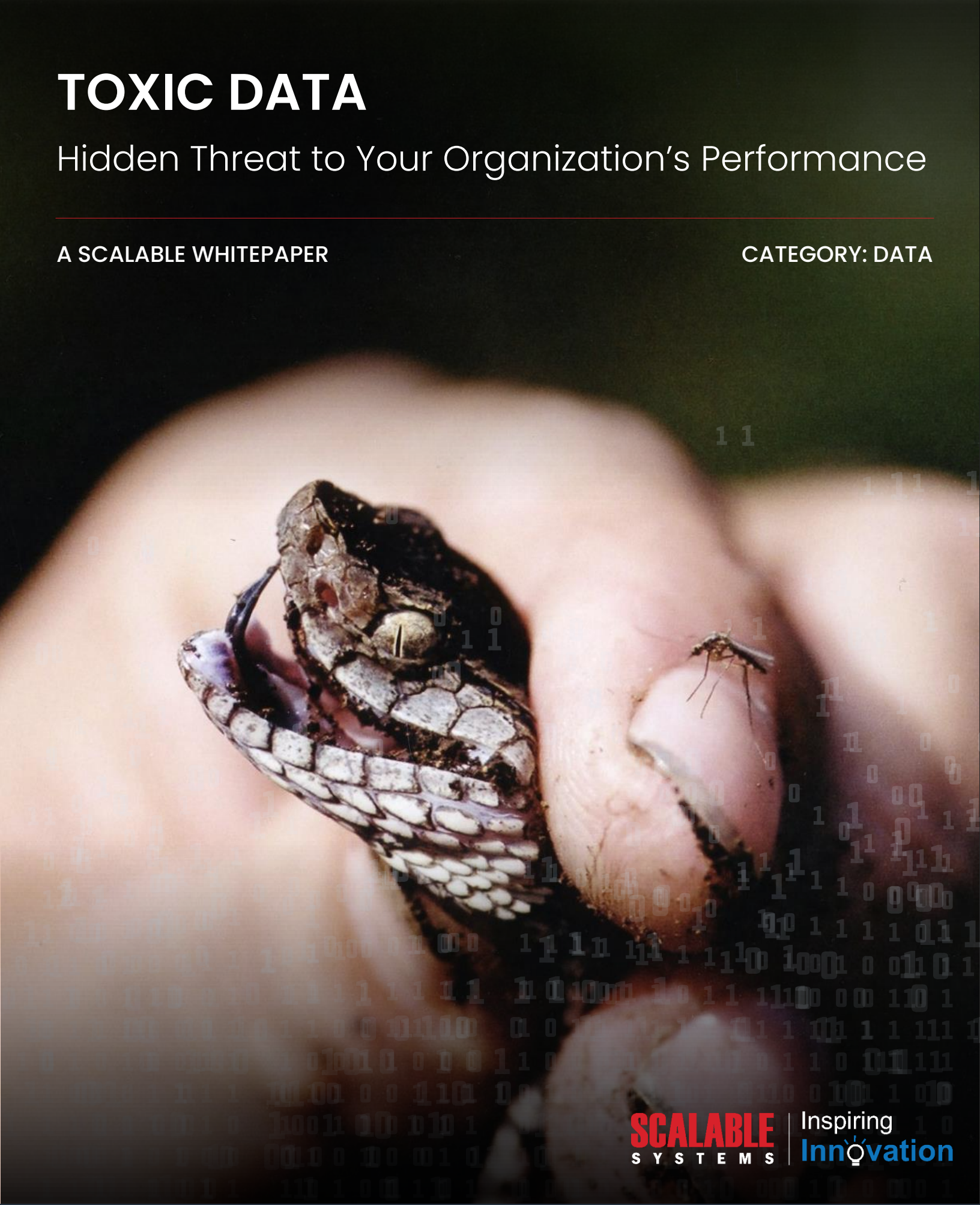
TOXIC DATA

Hidden Threat to Your Organization's Performance

A SCALABLE WHITEPAPER

CATEGORY: DATA



SCALABLE SYSTEMS | Inspiring Innovation

TABLE OF CONTENT

INTRODUCTION

Business of all sizes is experiencing a massive explosion in the volume of data. Though Data is recognized as one of the most important assets of the company, many times, Data is overlooked. As a result, data quality goes from bad to worse, making it toxic. Toxic data could be damaging to any organization. Data toxicity may not often be recognized but works as a silent killer damaging to business in various ways. Over time, the ongoing harmful intrusion of toxic data can make an organization sick. Suppose tainted data is not cleaned and prevented. In that case, it may lead to various organizational hazards like declining sales, poor customer service, inadequate response to customer demands, and inability to innovate and sustain in a difficult business environment. In earlier days, almost any business could survive if they focused on quality customer services, good product or service lines, and reasonable marketing efforts. Today even the best companies with exceptional customer services, innovative products or services, highly experienced management, and a dedicated team of employees cannot be guaranteed success due to global competitiveness, aggressive price reduction, easy information availability, and bad market condition. There are numerous exceptions to every business rule defined in past decades.

Data is the lifeblood of any organization. There should be a healthy data flow to maintain a healthy and happy organization. Toxic data at any department of the organization over some time may lead to data plaques which hampers the smooth flow of data. When a smooth flow of data hampers and data becomes toxic, here is a typical scenario in an infected organization. The operational data management team works harder to meet the data demand. Data pressure increases from the marketing team since customers are leaving. The finance department feels dizzy with regular short of quarterly targets. IT department experiences fatigue while doing everything they can but needs more results. Organization Management experiences increased sweating and cancel golf meetings. Suddenly organizational vision gets blurred, and the negative spiral continues.

Had the organization taken the preventive steps to keep the healthy data flow by Scalable Data Quality framework, regular data governance exercises, frequent data quality checks, identifying and preventing toxic data entry points, and data cleansing, it could have positioned itself as a healthy, strong, growing organization.

Bad Data can be very costly, particularly for small and medium-sized businesses where the difference between survival and closure can rest on the ability to recover from a disaster. At the very least, critical data loss will financially impact companies of all sizes. The economic implications for a company are a combination of loss of business, low productivity, legal action, and the cost of re-creating data. At its very worst, critical data loss can lead to business collapse.

OVERVIEW

According to Gartner Inc., organizations may lose more money in operational inefficiency due to data quality issues that they spend on Data warehousing and CRM activities. Data quality is important to get the desired business result for every organization. Though most organizations often recognize the importance of data quality, data quality initiative gets lost while coping with changes, planning for growth, and resolving daily operational challenges. Most organizations agree that they need to pay more attention to data while developing operating systems. Many Organizations plan to improve data quality by using various methods such as data profiling or data cleansing tools to cleanse the toxic dirty data with Extract-Transform-Load (ETL) tools for Data Warehouses and other important applications. All these technology-oriented data quality efforts are steps in the right direction. However, technology solutions alone cannot eradicate the root causes of poor-quality data because it is not as much an IT problem as a business one. Most of what is stated is obvious and may also relate to the incidents they must have faced in real-life.

Toxic data is a serious concern in today's business environment. Data quality issues must be addressed systematically and organizationally. Enterprise-wide data quality discipline must be established and constantly nurtured. Organizational data should be valued and treated as assets as other tangible assets like buildings, employees, and customers are treated.

HOW TOXIC DATA HAPPENS

Data Quality becomes a problem when organizations need to treat information as an asset that can bring measurable value for growth and profit. When data is not cleaned, scrubbed, checked, validated, measured, or not cared it gets contaminated. The same data having many duplicate versions, especially from legacy systems, leads to multiplicity syndrome creating great confusion and inaccuracies. Data Quality will be fine if there are a few data entry points and sporadic data usage. But "the environment doesn't sit still" when external applications such as ERP and CRM run in an organization, and it is very difficult to enforce data quality standards. When data is not complete, correct, and consistent, businesses often blame IT. It is natural for organizations to think of data as being IT's responsibility. However, IT departments cannot manage data quality by themselves. According to a study published by the Data Warehousing Institute entitled "Taking Data Quality to the Enterprise through Data Governance," data quality is mostly related to business issues.

DATA CAN BECOME TOXIC OVER SOME TIME DUE FOLLOWING REASONS:

POOR DATA ENTRY HABITS WITHOUT ADEQUATE VALIDATION AND QUALITY CHECK

Many legacy systems developed years back need more validation and checks to prevent data entry errors and anomalies. Also, if there is some validation data entry, the operator has often found easier ways to override it. For example, suppose a telephone number entry has a validation that the telephone number should have xxx-xx-xxxx format. In that case, an operator can easily override the validation by entering 111-11-1111, which is of no value.

LACK OF CLARITY IN BUSINESS RULES DEFINITION

If business requirements are not articulated precisely, that leads to speculation, spiraling down a wrong path of incorrect data modeling and great applications that brings process and reports erroneous data.

IMPROPER INTERPRETATION OF BUSINESS RULES

Business users involved in intermittent data entry activities often might be clear about some business rules. In that case, they might enter the data which they think is correct. This leads to data inconsistency and, if not resolved, can become toxic data.

POOR DATA CAPTURE

During system requirements definition, we rarely bother to gather the data requirements from downstream information of consumers, such as from the marketing department. For example, let's build a system for the lending department of a financial institution. The users of that department will most likely list the Initial Loan Amount, Monthly Payment Amount, and Loan Interest Rate as some of the most critical data elements. However, the most important data elements for marketing department users are probably the Gender Code, Customer Age, or Zip Code of the borrower. Thus, in a system built for the lending department, data elements such as Gender Code, Customer Age, and Zip Code might not be captured or haphazardly. This is often why so many data elements in operational systems have missing or default values.

POOR DATA MODELING AND DATA ARCHITECTURE

Data Modeling and Architecture need to be designed in a scalable way so that when the data size grows, and new applications are added, it should withstand the pressure of changes. The poor architecture will lead to duplicate entries, redundant data all across the systems, and improper correlation between tables. Eventually, as the load increases, the methods may collapse without warning.

IMPROPER DATA MAPPING FROM OTHER ERP AND CRM SYSTEMS

In a complex business environment, a continuous data flow flows from one system to another. If improper data mapping remains undetected, then toxic data starts circulating throughout the veins of the organization's data center. Often these errors are hard to detect because data mapping between heterogeneous systems often focuses on field type matching and standard validation. For example, if the first name in one ERP system is mapped to the last name in CRM systems, it might pass through since both are string values. This toxic data harms customer dissatisfaction, whose name is always spelled wrong when they get a letter from the company.

TECHNICAL ERRORS THAT OCCUR DURING THE TRANSMISSION OF DATA

Along with the growth of e-commerce, there came an increasing dependence on software programs to automate tasks involving databases of customer information. This opens doors for software programs to accidentally execute tasks that affect thousands or millions of records at a time incorrectly.

DATA ERRORS DURING APPLICATION MIGRATION

When applications are upgraded to other platforms for better performance and better user interfaces during the application migration, the application code used to handle data in a specific way might not take the same data element in the same manner after migration. So it is important to give special attention to data-sensitive applications.

DATA ERRORS DURING DATABASE UPGRADES AND UPDATES

When the database is upgraded, the new one may not support some old functions for calculating data values. The application might not handle the data as it should after the migration. If undetected in the production environment, the application will start bringing bad data, which can become toxic over time.

DATA QUALITY AS A NON-PRIORITY ISSUE

Many companies realized they needed to pay more attention to data while developing systems during the last few decades. While delivery schedules have been shrinking, project scopes have been increasing, and companies need help implementing applications in a time frame acceptable to their business community. Because a day has only 24 hours, something has to give, and what usually gives is quality, especially data quality.

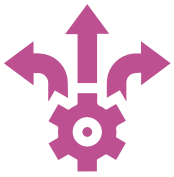
EFFECT OF TOXIC DATA ON BUSINESS

The consequences of toxic data quality are real. At the most basic level, tainted data can affect revenue, costs, and Customer loyalty. Data quality is a critical component of business success. Poor quality data jeopardize the performance and efficiency of operational systems. It also undermines the value of business intelligence systems on which organizations rely to make key decisions. Decisions based on such data can cause direct financial loss, spoil customer relations, and damage an organization's credibility in the marketplace. As more organizations recognize data as a strategic asset, business leaders are increasingly held accountable for ensuring information accuracy, quality, and reliability.

Poor data quality adversely affects your organization in three keyways:

POOR DATA QUALITY CAUSES INEFFICIENCIES IN THOSE BUSINESS PROCESSES WHICH DEPEND ON DATA

Almost every business process is dependent on data in some way or another. From customer order entry, invoicing, and reporting, business analytics data plays an important role. Even if all business processes are near perfect, poor data will change everything.



POOR DATA QUALITY GIVES RISE TO POOR DECISIONS

A decision can be no better than the information upon which it's based, and critical decisions based on poor-quality data can have very serious consequences.

POOR DATA QUALITY CREATES MISTRUST

Time, money, and reputation can be lost if the data's wrong. According to some studies, the cost of losing one customer is four times higher than obtaining that customer due to advertising costs and marketing staff expenses.

PROCESSES THAT TOXIC DATA CAN IMPACT

HOW TO DETOXIFY TOXIC DATA

For detoxification, the existing data needs to be analyzed carefully and scalable. There might be lots of places the data quality could be better. But all data might be cleaned after some time. Focus on the immediate problems which poor data might cause. Once the problem area is identified, the following methods can be used to analyze, identify, and clean the data.

DATA PROFILING

Data Profiling is a process where data content, structure, and relationship are evaluated and measured. Data anomalies like empty columns, new data values, overused data values, duplicate data columns, violation of structure rules, business data rules, and representation of missing values are discovered during data profiling. Many popular data profiling tools or in-house by SQL queries can do data profiling.

DATA CLEANSING AND ENHANCEMENT

Data Cleansing is a process of correcting or removing toxic data. Data cleansing is a repeating process till all data issues are cleared. Data cleansing requires a thorough understanding of the business objective. Data cleansing involves looking for and handling data errors, outliers, and missing values.

DATA MATCHING AND CONSOLIDATION

Match similar records and perform de-duplication and consolidation based on set criteria. Matching is done on various business rules like name, address, SSN, and DUNS. Once the duplication is determined, the merge is performed on duplicate records if they are the same identity. This directly benefits the removal of a duplicate from your database and Improves the accuracy of Customer Information in the Customer database.

HOW TO PREVENT TOXIC DATA AND MANAGE DATA QUALITY

90% of the toxic data enters an organization's various data entry points. If bad data entry can be controlled at the source checkpoint, the organization can save a significant amount later on data correction and detoxification. Entering your data correctly for the first time is the best way to ensure the integrity of your data. Be prepared to spend money at this data collection stage. It saves more money in the long term. Use auto-complete or other data validation applications at the data entry data stage. Using controls and input masks during data entry can help to correct entries in formatted fields.

A data quality framework should be established and followed for ongoing data quality improvement.
The data quality framework is intended to provide a common objective approach to assessing data quality. There are six data quality dimensions.

COMPLETENESS Is your data complete	**CONFORMITY** Does your data conform to out standard	**CONSISTENCY** Is your data consistent
ACCURACY Is your data accurate	**DUPLICATE** Is your data containing lots of duplicate	**INTEGRITY** Is your data maintaining integrity

Data quality is an iterative process of assessment, planning, and implementation. Iteratively repeat the data quality process to measure ongoing effort and effectiveness. Assessment results can be used to build an economic model that evaluates the costs of instituting improvements. This model can be viewed as a scorecard that documents data quality levels associated with data quality dimensions measured at specific locations in the information chain. Here are the steps involved in building a data quality scorecard to summarize the overall cost associated with low data quality and help identify the best opportunities for improvement:

The data quality scorecard is a framework for calculating the return on investment for improved project implementation. The scorecard can be used as a management tool, in which any suggested improvement is connected with the cost of designing and implementing the advance, along with a time frame for implementation. Ultimately, this scorecard can be used as the basis for an ongoing data quality improvement project to enhance the company's intelligence efforts.

Data Management

Aligning your business's data strategies with key data management enablers

Data Quality Management

Fuel trusted insight generation and evidence-based decision making

ECM Lifestyle Management

Manage your organization's content and documents across their lifecycle

Reporting, Analytics & AI

Manage your master data, build out that single stakeholder view

DATA PLATFORM & PRACTICE

Build the right platform to deliver data from source to analytics

UNLOCK THE VALUE OF DATA

Data Strategy

Treat data as the corporate asset that it is

Data Governance

Establish and enabling framework that drives use of and appetite for data

Enterprise Content Management (ECM) Strategy

Optimize searchability and findability of information assets

Data Integration

Manage your master data, build out that single stakeholder view

CONCLUSION

Data quality is an ongoing process and cannot be achieved overnight. As per the Japanese Kaizen principle, small daily improvements eventually result in huge advantages. Data quality is a broad umbrella term for the accuracy, completeness, consistency, conformity, and timeliness of a particular piece or set of data and how data stores and flows through the enterprise. Different organizations will have different definitions and requirements for data quality, but it ultimately boils down to data “fit for purpose”. Data as an asset must be usable by the organization for constant growth.

We at Scalable Systems view our solutions approach to customer data as an art form - because it is both a creative and constantly evolving process. Rather than merely cleansing and organizing your database, our preference is to continually nurture, manage, cherish, and maintain your data to ensure it does not become toxic at any point now or in the future. With our expertise in Data Model architecture, Database Administration, Data migration, sound database development, Data quality framework, and Master Data Management, we provide holistic and long-term solutions for the most important asset of your organization – Data.

About Scalable Systems

Scalable Systems is a Data, Analytics & AI Company focused on vertical-specific innovative solutions. By providing next-generation technology solutions and services, we help organizations to identify risks & opportunities, and achieve sales and operational excellence to gain an innovative edge.

www.scalable-systems.com