WHITEPAPER

An Approach to Ensuring Good Data Quality for Enterprise Data

Data quality management strategies for a resilient data mature future

Gleecus TECHLABS INC

Forward Thinking Digital Innovation

Table Of Contents

Executive Summary

Good quality data is imperative to the success of analytics and AI initiatives of any organization. Organizations heavily rely on these technologies for their business decision making and innovation at the same time, they are handling multiple data flows in humungous quantity that can leave them overwhelmed and unable to maintain quality of data.

Poor data quality makes the organization vulnerable to several risks. It leads to reduced efficiency, missed opportunities, reduced revenue, and impaired decision making. Wrong reports and analytics from bad quality data also causes the trust in analytics to dissolve for employees, customers, and business leaders.

This white paper takes a high-level view of the perils of bad quality data and the possible solutions to escape. It outlines various strategies, tools, and practices that can improve the capability of your organization in maintaining the quality of data and moving towards becoming a truly data-driven business.

The Imperative of Quality Data

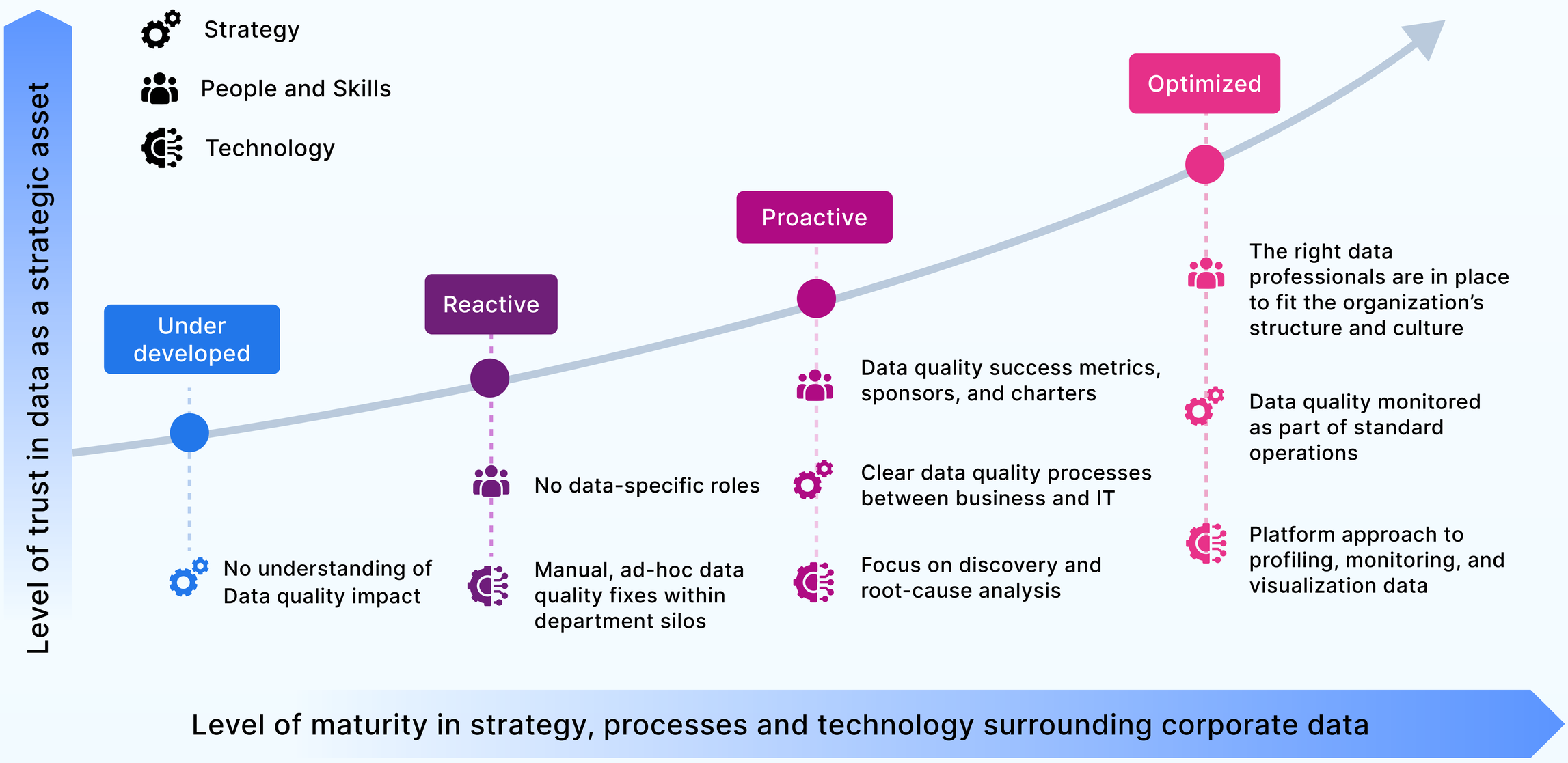
The Changing Data Landscape

Today data is generated more than organizations realistically can handle without any special strategy. Organizations are harvesting data from several data sources simultaneously like Customer Support Channels, Enterprise Knowledge Centers, ERP, CRM, Human Capital Management (HCM) systems, Financial systems (accounting, budgeting), Machine-Generated Data, and millions of other datapoints. They highly rely on data for their business decisions. This makes quality data imperative to the success of insight-driven businesses. The velocity, volume, and variety of data generated can result in inaccuracies in data and data handling/manipulating issues without proper ways to check the state of data quality. Without data-skilled professionals or a system to verify data, organizations can be challenged with unreliable data, leaving business decisions to erroneous analytics.

How Organizations Are Responding to the Changing Data Landscape

Organizations are heavily investing in automation, observability, data skills, and agility to uplift their data quality maintenance processes and technology framework. It is a continuous process of evaluating their position with respect to a data quality maturity model and move from a reactive stage of ad-hoc data quality fixes to an optimized stage to a platform approach to profiling, monitoring, and visualizing data.

The Data Quality Maturity Curve



Over three quarters of leaders say that data quality automation will be important to their organization moving forward. They are investing to improve Data Observability, a company's ability to thoroughly understand the health of its data. This ensures businesses are consistently using valid information for reporting and decision-making by conducting regular data health checks across all systems and procedures. Following are some of the objectives organizations trying to fulfill through their data initiatives:

Improving and accelerating decision-making

Better and faster decisions using data is a top priority to respond to market pressures. A continuous influx of accurate data enables team members—technical or not—to act with confidence. Quality actionable data is vital in a market that is moving faster than ever.

Increasing agility to better manage uncertainty

60% of business leaders agree that agility is an important priority to better manage uncertainty. In recent years, agility has become increasingly important as the pandemic and economic distress reshaped our lives. Further, a recent research study indicated that agility is a top benefit of data quality automation, which is a key differentiator between super performers—those who have exceeded annual goals and targets even while responding to market obstacles—and everyone else at a lower stage of data maturity.

Scaling business growth

Businesses are not looking solely to survive turbulent times; they must thrive to stay relevant in today’s marketplace. With increased competition, it’s important for businesses to have the most up-to-date consumer information to better understand who their target audience is, what they care about, and how to reach them effectively. Growth starts with a data-driven foundation.

Investing in the digital experience

Enterprises are prioritizing digitizing the customer experience as they respond to these market pressures. In today's time a digital omnichannel experience is nonnegotiable. The impact of data quality is two-fold as it helps improve the experience of online applications and collects high volumes of data for further analysis and reporting.

Automating data quality processes to reduce cost and improve efficiency

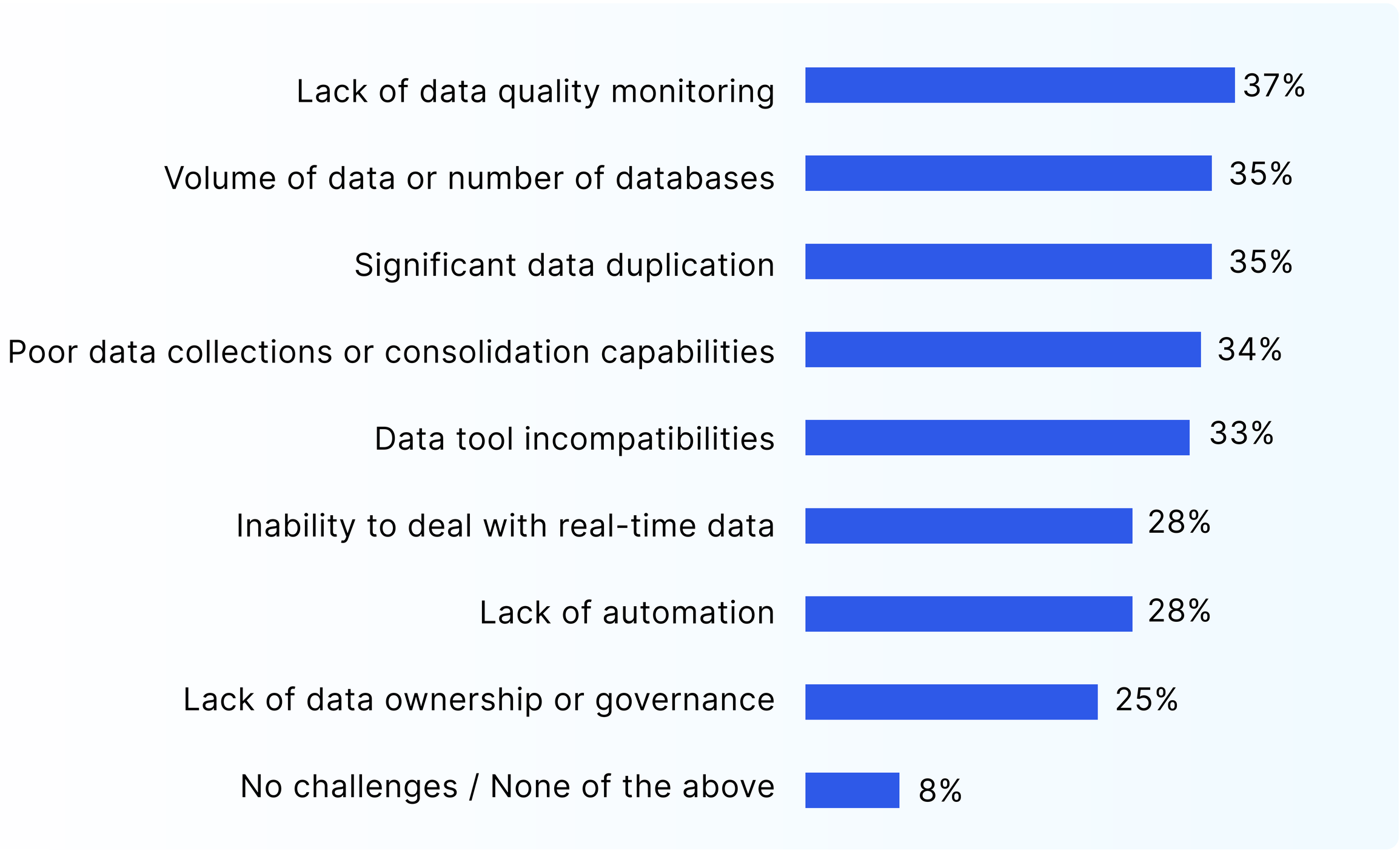
Business leaders are finding value in automating data quality processes to reduce costs and improve efficiency. Data quality automation has become a hot topic over the last several years as it helps address various obstacles from skill shortages to agile workflows and digital acceleration. Data quality is the common thread that supports these priorities.

The State of Quality Data

> *According to a survey 92% of leaders experience challenges with maintaining good quality data.*

Without a proper data governance and data quality monitoring system in place most organizations fail to manage countless streams of data. Lack of data governance in many organizations implies data ownership, management, and control is fragmented with siloed and decentralized data ownership. While organizations acknowledge the benefits of utilizing their enterprise data and accept there is a need to improve quality of data, they find a hard time to start. They are riddled with questions like: How to define what high quality data means and how to measure it? How to modernize the existing data infrastructure to ensure high data quality?

Top Challenges of Maintaining Good Quality Data



**According to a survey of over 500 people within US organizations across a variety of industries and roles by Experian*

How Poor-Quality Data Poses Business Challenge

Poor quality data is seen to impact the organization by compromising business agility and the ability to adapt quickly to changing market conditions, inhibiting the ability to understand customers and deliver a personalized customer experience, wasting resources and escalating costs, damaging the reliability / trust in analytics. Bad quality data can lead to incorrect insights, poor decision-making, and costly mistakes. Here are a few examples:

Medical Misdiagnosis: Imagine a hospital relying on outdated patient records. A patient could receive the wrong treatment based on inaccurate data, potentially endangering their life.

Incorrect Financial Forecasting: Banks relying on inconsistent financial data might make incorrect predictions, affecting everything from loan approvals to stock investments, ultimately impacting their bottom line.

Manufacturing Defects: A manufacturing firm using incorrect data about material properties or machine calibration settings might produce defective products, leading to costly recalls and damage to the brand's reputation.

Marketing Campaign Failures: A marketing team working with bad customer data might target the wrong audience, leading to poor campaign performance and wasted advertising budget. A notable incident occurred with a major retailer that mistakenly targeted a large portion of their customer base with promotions for products they had no interest in, resulting in low engagement and wasted resources.

Poor Inventory Visibility: A retail company using erroneous inventory data might overstock unpopular items while running out of bestsellers, leading to lost sales and unhappy customers.

Poor Energy Audit: An energy company using inaccurate consumption data might overestimate demand, leading to excess production and wasted resources. Conversely, underestimating demand could result in blackouts and customer dissatisfaction.

Transportation and Logistics Delays: A logistics company using faulty data for route planning might experience delays, increased fuel consumption, and higher operational costs. For example, a leading shipping company once faced significant delays and increased costs due to relying on outdated and incorrect traffic data, leading to missed deliveries and unhappy customers.

Government Policy Missteps: Governments relying on inaccurate census data might misallocate resources or design ineffective policies. For instance, a city government once used outdated population data to plan public transportation routes, leading to overcrowded buses in some areas and underutilized services in others, causing public dissatisfaction and inefficiency.

CASE STUDY

WELLS FARGO

A well-known example of bad data quality leading to severe consequences is the Wells Fargo account fraud scandal. Employees, under pressure to meet sales targets, created millions of fake accounts. Due to poor data quality and oversight, these fraudulent activities went unnoticed for years. The fallout included substantial fines, loss of customer trust, and severe damage to the company's reputation. This scandal highlighted the critical need for accurate data and robust data governance to prevent misuse and ensure transparency.

Establishing Proper Data Hygiene

Determine the referential value of your data with profiling	Data profiling is the process of gaining an understanding of the existing data relative to the quality specifications. Audit and enumerate your existing datasets with a business objective, defining, usage and goals.
Establish clear data ownership	Use the metadata of the data repository to profile how and where organizational data is store, and who is responsible for the repository.
Avoid getting complacent	Enact a data custodian role responsible for documenting and preserving the findings regarding data sources and processes.
Avoid the "P-Hacking" tendency	P-Hacking is the concept of using inaccurate or purposedly manipulated data to achieve the desired results. Do due diligence to analyze, evaluate, and ensure the quality of data.
Lead with mature Data-Ready managers	Induct a dedicated team of data professionals who pave the way for a data driven culture and business users with access to insights and the knowledge to manage the information appropriately.

Data Quality Management

With so much data at our fingertips, businesses that can sustain success in an unstable situation are often the ones who are investing in data management and reaching for optimal data maturity. Different data management methods have been designed to help scale the quality and use of information. As data observability becomes more prevalent, organizations see the value of proactively cleaning their data for actionable insights.

What is Data Quality Management?

Data Quality Management (DQM) involves all activities that are related to the conceptual structure of an organization's data. This includes strategic and organizational aspects such as technological decision-making and the selection of appropriate methods to improve the overall quality of data. It involves a virtuous cycle of ongoing observation, analysis and improvement of data. This enables organizations to be proactive in controlling the health of their data, as opposed to fixing erroneous data after it’s been identified and dealing with its consequences.

The Need of Data Quality Management Tools

Organizations thrive on optimizing their operations and maximizing their return on investment (ROI). To achieve these positive outcomes, organizations rely heavily on data-driven insights and processes to instill trust in their business decisions. Enabled through DQM, high-quality data and analysis allow businesses to capitalize on opportunity, attain a competitive advantage and identify when corrective actions are needed.

For these reasons, organizations choose to invest in DQM software and tools with following objectives:

Building a foundation for business initiatives

Inaccurate data leads to increased business rework and mistakes. Implementing a DQM process assists companies in defining a framework for each department to follow and ensure data quality upkeep.

Remaining ahead of trends

Good quality data plays a fundamental role in allowing businesses to stay relevant and avoiding becoming a ‘dinosaur’. A recent study shows 79% of enterprise executives believe that failing to embrace big data will harm their competitive position.

Supporting successful data migrations.

Bad quality data is one of the key reasons why data migration projects fail. To safeguard the success of these projects, data rules should be introduced to identify and fix errors before a migration begins.

Minimizing processing time and costs.

Dirty data can be very costly to a business. Not only does it lead to wasted time reworking information, but it also leads to costly errors, such as losing track of orders or business expenses.

Meeting compliance and risk objectives.

DQM is an effective way to ensure organizations are operating correctly and are using information in line with data regulations.

Data Quality Management Methods

Data Profiling

Effective DQM software should include data profiling features. This refers to the process of collecting and organizing information so that it can be more easily examined, uncovering its maximum value. Data profiling automates the process of identifying and evaluating data to collect information about the data set's quality and hygiene. This provides critical insights for the business and improved visibility into any data quality issues, risks and overall trends.

Data Remediation

Data remediation involves data cleansing the process of correcting or removing corrupt records, such as duplicate, inaccurate or redundant data before it is loaded onto a destination. After data has been cleansed, it can be enriched. This involves the process of merging data from an external authoritative source to enhance current data to make it more useful. Modern data cleansing involves utilization of automation tools like ETL (Extract, Transform, Load) that takes care of extracting, cleaning, and converting data into standardized format.

Data Lineage Management

Tracing the source of data can be an onerous task. Data lineage refers to the ability to trace the origin of data and its movement over time. This helps control the flow of information and can provide visibility into the root cause of data discrepancies.

Data Quality Checks

To protect the integrity of a database, each incoming record needs to be tested against pre-established data quality dimensions. DQM checks ensure that rules are integrated into the flow of information, allowing for the monitoring and reporting of any data errors.

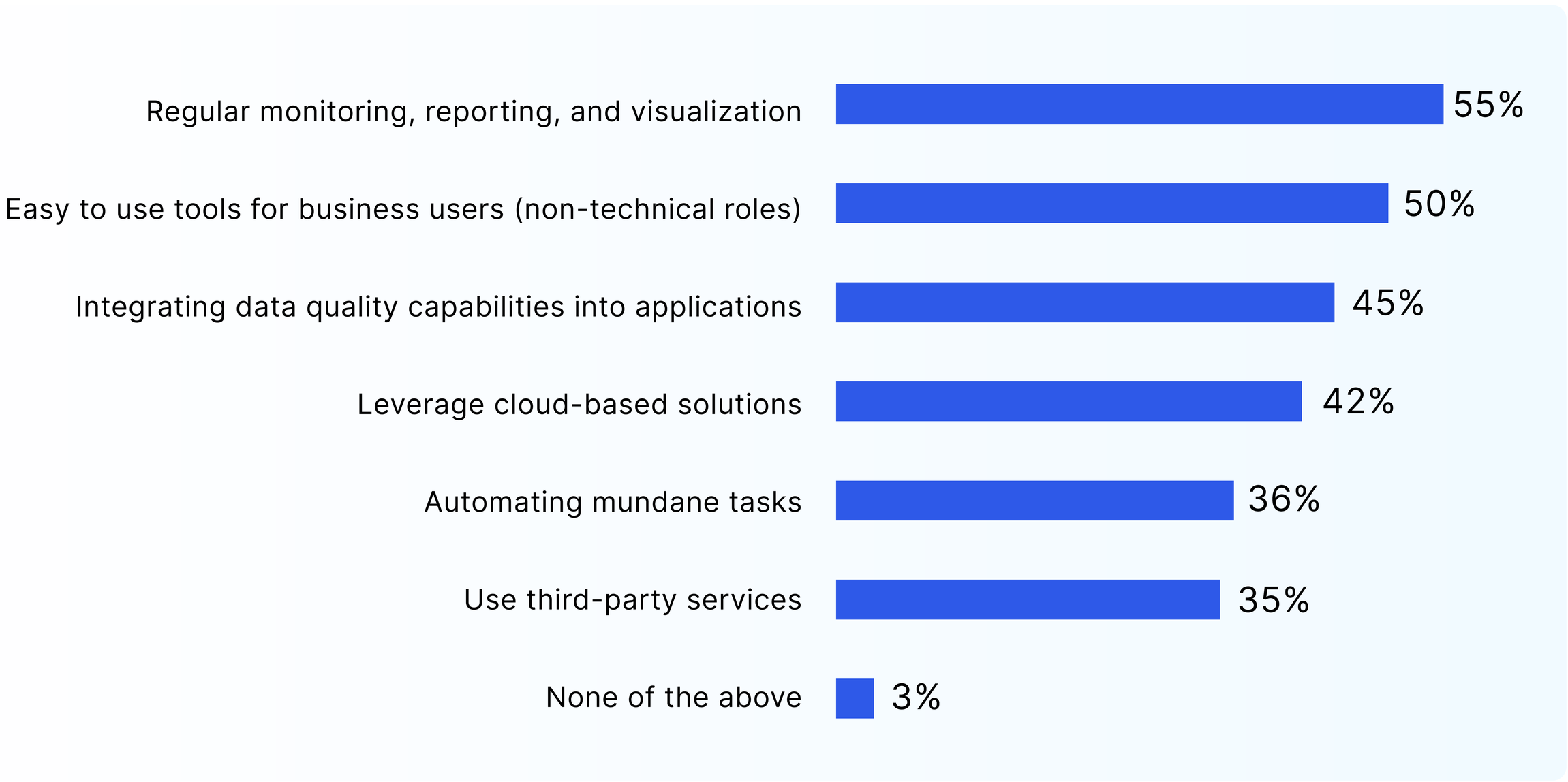
Data normalization

Data Normalization is a way to check redundancy or chances of data duplication by organizing and standardizing similar data constructs. For example, if the business rule is that all dates should be mentioned in the MM/DD/YYYY format, data cleansing can use pattern matching to find the inaccurate types and replace them with the correct data format.

Data augmentation

Data augmentation is the process of adding something to an incomplete dataset so that it conforms to the standard schema. This process improves the quality of data and makes it fit for use.

Data Quality Solutions



**According to a Experian survey of over 500 business leaders and managers sitting in departments like IT, data, operations, finance, customer service, and others across a variety of industries.*

Data Quality Tools and Platforms

Many organizations struggle to keep up with the demand of DQM procedures, despite how valuable or fundamental they are to the business. Often, they find they are lacking the internal resources, such as personal capabilities or outdated technology, to manage the volume and complexity of data effectively.

Competitive organizations must establish reliable data management measures to maintain the usability and value of their data. As DQM is a complicated affair, companies have to make the choice of handling it in-house, outsourcing consultants or a combination of both.

- **Automation of data quality rules and processing**
- **Data quality CRM/ERP/CDO or other data quality integrations**
- **Stand-alone data quality platform**
- **Data cleansing and quality capabilities built into other software tools**
- **Unified data management platform including data catalogue and governance**
- **Contact data validation solutions**

Here are some popular DQM tools and their contributions:

Talend Data Quality: Ensures data integrity and quality through real-time data profiling and cleansing capabilities. Specialty: Data profiling, cleansing, and enrichment.

Informatica Data Quality: Automates data cleansing, monitoring, and profiling across the organization, improving overall data quality. Specialty: Comprehensive data governance and integration.

IBM InfoSphere QualityStage: Focuses on data cleansing and entity resolution to ensure high data accuracy and consistency. Specialty: Enterprise-grade data cleansing and matching.

AWS Glue: An ETL (Extract, Transform, Load) service that helps ensure data is well-structured and ready for analytics. Specialty: data cataloging, schema discovery, and data transformation.

Roles Ensuring Data Quality and Maturity

> *95% of organizations are hiring or looking to hire data roles in the near term.*

Understanding the roles organizations are investing in can help identify their strategic focus. As DQM is a resource-intensive process, it is important businesses delegate the right employees to specific and well-defined roles. This will involve cross-departmental cooperation and alignment with C-level executives.

Data Owner

Responsible for monitoring the appropriate access to data and also controlling the risk. They are the central contact people and have the right to edit data and how it is used.

Data Quality Analyst

Examines and summarizes data. Conducts data profiling and assembles the results for presentation and review with stakeholders.

Data Steward

Concerned with the meaning of data and ensuring that the data is correctly used. Responsible for knowing how the data is collected, maintained and interpreted.

Data Custodian

Responsible for managing the actual data, this includes handling servers, backups, or networks.

Metrics That Define Good Quality Data

Ensuring DQM involves validating the accuracy, completeness, consistency, freshness, and relevance of data, which forms the bedrock of trustworthy analytics.

Accuracy

Accuracy means that a measured value accurately reflects the true value, free from errors like outdated information, redundancies, and typos. To ensure the highest quality of your data, focus on continually improving its accuracy as your datasets continue to grow in volume.

Completeness

Ensures data records are complete and contain enough information for accurate analysis, it’s important to track this data quality metric by identifying any fields with missing or incomplete values. Comprehensive data entries are essential for maintaining a high-quality dataset.

Consistency

This implies data in your databases should be free from contradictions, ensuring that values from different datasets align correctly. To verify consistency, you might establish and adhere to specific data rules.

Freshness

Freshness measures the delay between when an event occurs and when it is recorded. This metric ensures that your data remains accurate, accessible, and available for your analytics and AI initiatives.

Validity

This ensures the consistency and correctness in format of data aligned to the established formatting rules. Tracking validity involves the percentage of format error for a data item against its total number of appearances across the database.

About Us

Gleecus TechLabs Inc. is one of the fastest growing IT innovation partners for startups, SMBs, and enterprises that help clients envision, build, and run more innovative and efficient businesses. We help enterprises in their journey of becoming a Managed Data Mature organization through innovative and trusted ways of observing, understanding, remediating, and controlling your data.

Our team of specialized data professionals expertly handles niche roles of data analyst, data stewards, data engineers and more to introduce clear ownership and transparency around your enterprise data. We build custom data engineering solutions and platforms for streamlined management of data and improved awareness and utilization of data.

Build a resilient system of data hygiene that fuels your journey towards a data mature organization

Connect with Us

About Gleecus TechLabs Inc.

Gleecus TechLabs Inc. is an ISO 9001:2015 and ISO/IEC 20000-1:2018 certified Forward Thinking Digital Innovation partner creating impactful business outcomes with Engineering & Experience. With deep focus on Cloud, Data, Product Engineering, AI and Talent we help organizations become Digital Natives.

Forward Thinking Digital Innovation

Email: hello@gleecus.com

Phone: +1 347 947 2022

www.gleecus.com