snapLogic®

RESEARCH REPORT

The State of Data Management

Why Data Warehouse Projects Fail

In part one of a two-part series, new research commissioned by SnapLogic and conducted by Vanson Bourne uncovers the challenges blocking data warehousing success

Contents

snaplogic

Introduction

Data is the fuel of our modern world, and its increased proliferation within organizations means that proper data management has never been more critical to success.

More than ever, organizations are investing in data warehouses and data lakes to help them make the most of their valuable data assets and deliver on the promise of agile analytics and actionable business insights. However, the process of identifying and moving data into a data warehouse or data lake is not always straightforward, all too often inhibiting progress and success.

Based on new research commissioned by SnapLogic and conducted by Vanson Bourne, who surveyed 500 IT Decision Makers (ITDMs) at medium and large enterprises across the US and UK, this whitepaper explores the data management challenges organizations are facing, the vital role data warehouses play, and the road to success.

Key findings

- **83%** of ITDMs are not completely satisfied with the performance and output of their data management and data warehousing solutions
- **88%** of organizations experience challenges trying to load data into data warehouses, with the biggest inhibitors being legacy technology, complex data types and formats, data silos, and data regulatory requirements
- The average organization reports losing over **$1 million USD** each year due to poor data management
- The average enterprise has **115** distinct applications and data sources, with almost half of them **(49%)** siloed and disconnected from one another
- **89%** of ITDMs are worried these data silos are holding them back
- ITDMs report that, on average, **42%** of data management processes that could be automated are currently being done manually, taking up valuable time and resources
- **93%** of ITDMs believe improvements are needed in how they collect, manage, store, and analyze data
- **76%** have increased their budgets for data management in the past year

Following the survey findings, the whitepaper explores the opportunity around data automation, and offers seven steps to help practitioners achieve data warehouse success.

The rise of the data-driven enterprise

The importance of data to organizations has grown exponentially in recent years, no longer prized only by data specialists but now a critical asset used on a regular basis by workers across each and every business function, from marketing to finance to HR.

It's therefore unsurprising that 98% of organizations surveyed as part of this study report that data is reviewed and analyzed on a weekly basis, underscoring the value of data-driven strategies and actionable insights across the enterprise.

Data is being collected, reviewed, and analyzed across all departments, from the top down. Over half (52%) of our survey respondents reported that those at the C-Suite and Managing Director level review data on a weekly basis (Fig.1), with data driving almost two-thirds (62%) of strategic decision making - a figure which is likely to increase as 70% want to achieve faster, more-informed decision making using data (Fig.2).

Who reviews data weekly within organizations

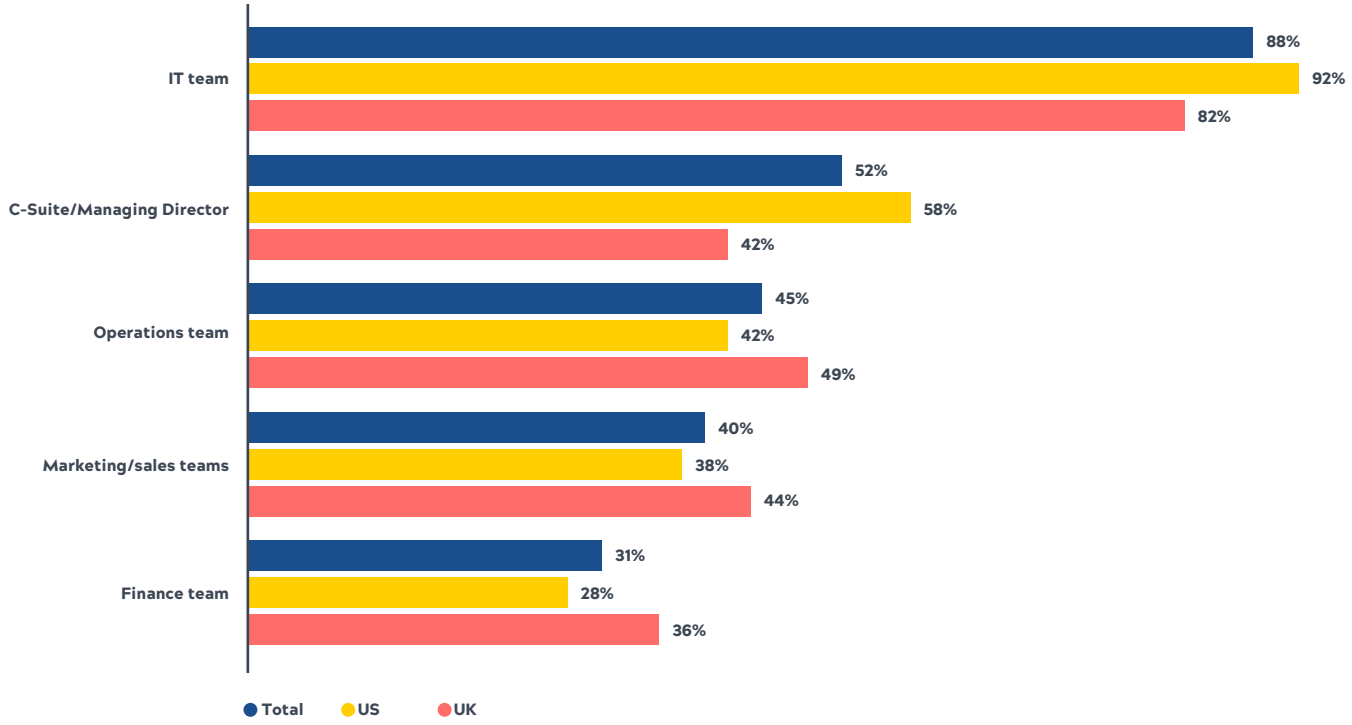

Figure 1: Analysis showing within organizations, which roles/departments review and analyze data on a weekly basis. Asked of all respondents [500] Split by respondent country

The value of data across the business

Faster and more informed decision making isn't the only benefit organizations hope to gain from better use of data (Fig.2). Notably, 64% want to use data to improve current business operations and 63% want to gain insights to drive innovation and business development. In the US, 65% are focused on using data to drive a better customer experience, while 60% expect data to enable employees to work with greater efficiency and agility.

Benefits from better data use

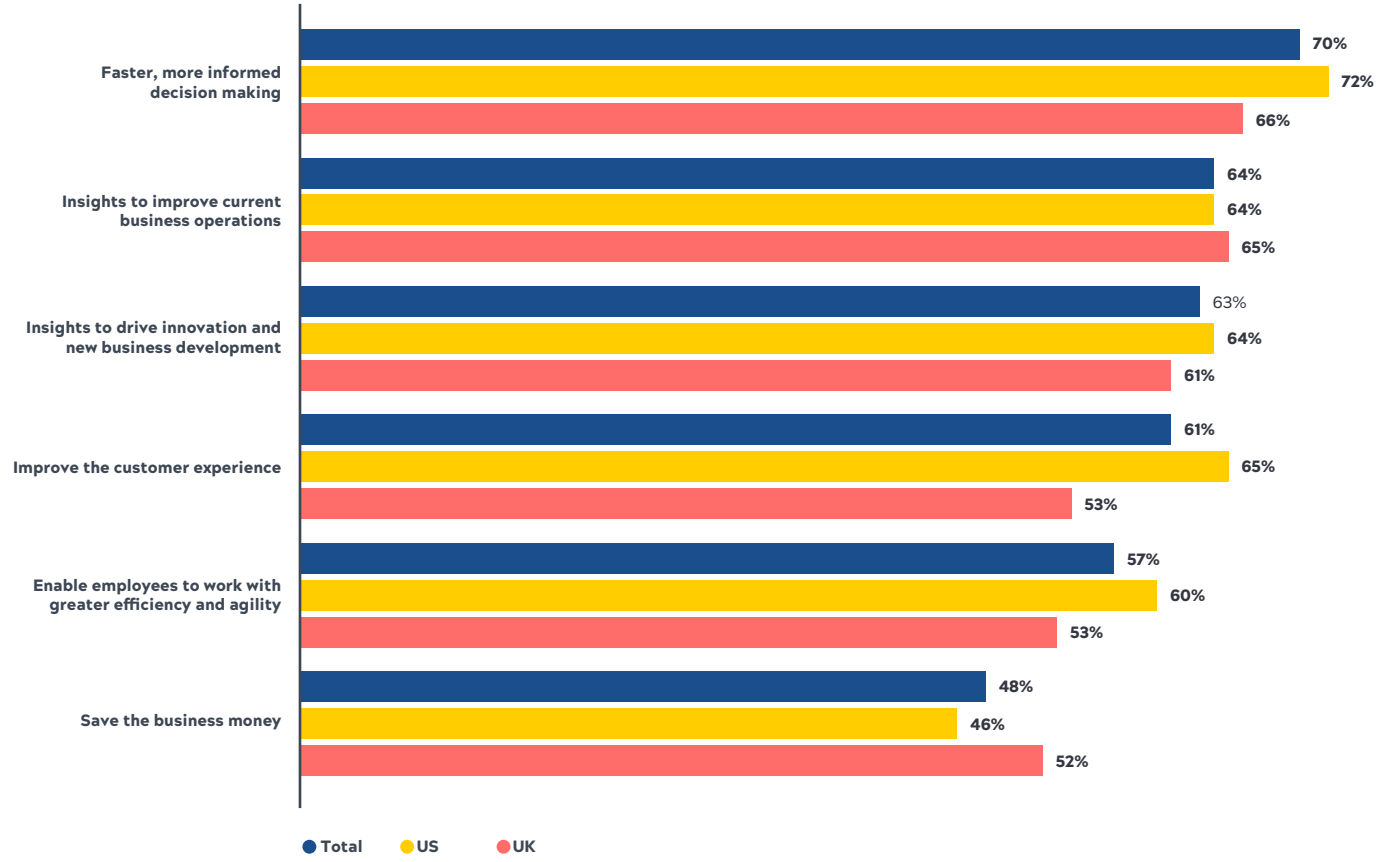

Figure 2: Analysis showing what organizations hope to achieve with the data they are collecting, managing, and analyzing. [500] Split by respondent country

For ITDMs, their organizational IT strategy will help to inform how they achieve these benefits. Most data management processes, from data acquisition through to analytics, are important to IT strategies (Fig.3), a recognition of the value that data serves to organizations.

IT strategy priorities

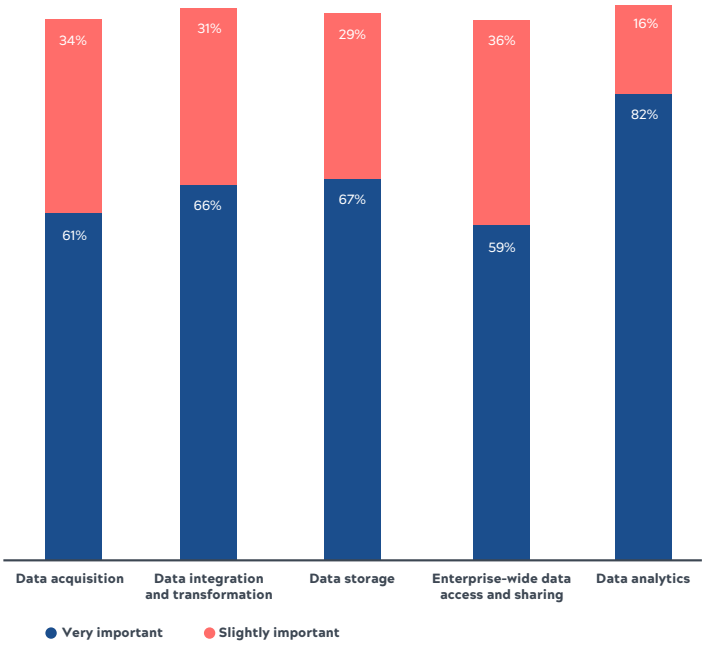

Figure 3: Proportion of respondents who report that the following areas are "Very important" or "Slightly important" to their organization's IT strategy [500]

Among these IT priorities, a slightly greater importance tends to be assigned to data analytics, with 82% saying it is "very important" to their strategy, perhaps not surprising as this is the final step in the process which directly produces the noticeable benefits drawn from data.

However, if organizations fall behind in the other areas of data management - including data acquisition, integration, transformation, storage, and so on - the data upon which analytics is based may not be timely or of good enough quality to provide the accurate insights organizations need.

To achieve its goals in these areas, an organization needs a variety of effective data management tools and processes. This ensures that all the data an organization possesses can be located, integrated, stored, and made available for analytics tasks when needed.

Increasing data management budgets

Notably, despite a highly uncertain economic period due to COVID-19, many IT leaders are continuing to allocate more budget for data projects than they have in the past. In fact, 73% of respondent organizations have increased budgets for data projects by up to 50% in the past year (Fig.4).

Rise in budget for data-related projects in the past year

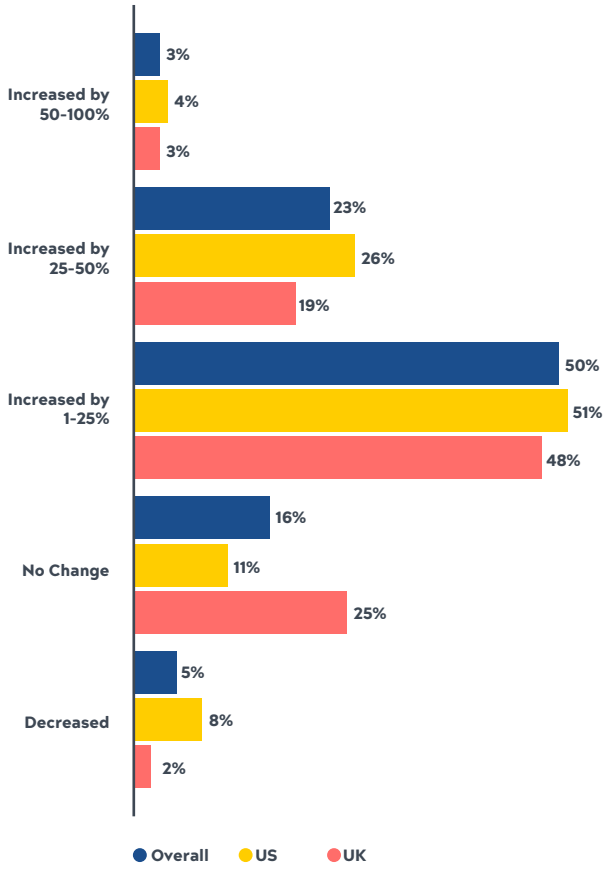

Figure 4: Rise in organizations' budget for data-related projects in the past year [500] Split by respondent country

snapLogic

Decisions to dedicate additional financial resources to data projects, in these uncertain times and in the face of competing demands for investment from other areas, are clearly indicative of the important role that data-related projects play - and will continue to play. Whether to fix or accelerate existing data projects, or to fund new data initiatives to help them emerge stronger on the other side of the pandemic, organizations are backing their commitment to data with the investments needed to make them a success.

But a well-funded and thoughtful data strategy isn't automatically a silver bullet. A lot rides on the insights that data can offer, and if this data is flawed or isn't able to be delivered for analysis in time, the risks for organizations are great. In fact, organizations, on average, report losing more than $1 million ($1.02m USD) each year (Fig.5) due to poor data management. For larger enterprises in the US, the projected loss from poor data management tops $1.5 million.

Financial impact of poor data management

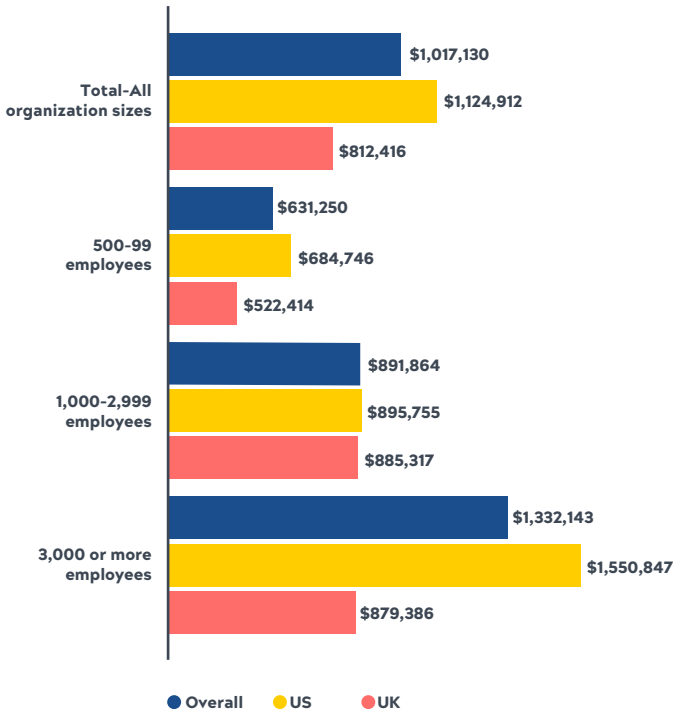

Figure 5: Average estimated annual financial cost (in lost time, resources, duplicated effort, missed opportunities etc.) to respondents' organizations as a result of poor data management (in USD) [500] Split by organization size and respondent country

Data management challenges

Effective data management is not without its challenges. Data can be hard to locate, or may be missing, or duplicated. Data formats and structures come in many forms and are not easy to reconcile. Preparing data for analysis can be time-intensive and resource-draining. All of these issues stand in the way of getting the fast, informed insights that organizations desire.

According to our survey, 83% of ITDMs are not completely satisfied with the performance and output of their data management and data warehousing solutions (Fig.6). This dissatisfaction stems from issues of data formatting (41%), data access issues due to regulations (33%), and speed of data movement and accessibility (31%) within the organization (Fig.7).

Data management satisfaction

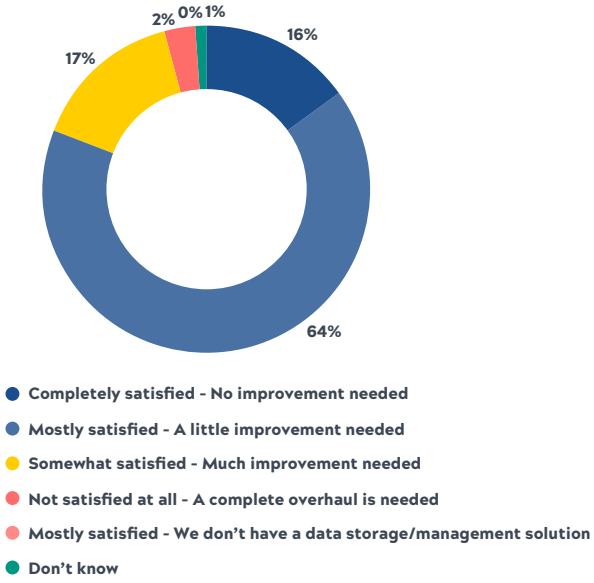

Figure 6: Satisfaction with the performance and output of their organization's data management/storage solution(s)

Reasons for data management dissatisfaction

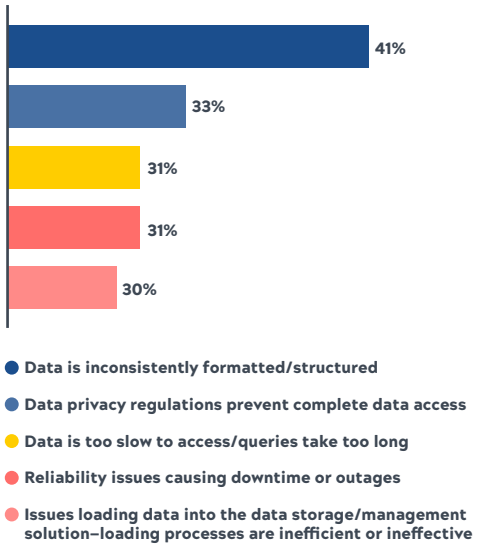

Figure 7: Analysis on why respondents are not completely satisfied with the performance and output of their organization's data management/storage solution(s). [414] Respondents who are not completely satisfied with their data management/storage solution, omitting some answers

These issues are only compounded by the growing number of applications and data sources organizations must now manage.

Disconnected data silos

Currently the majority of organizations (71%) are utilizing operational or departmental databases to meet their data storage and management needs (Fig.8), as opposed to an enterprise-wide data warehouse (55%). This raises some concerns given the propensity for these departmental databases to be managed as separate entities with a lack of integration between them.

Data storage/management solutions in use

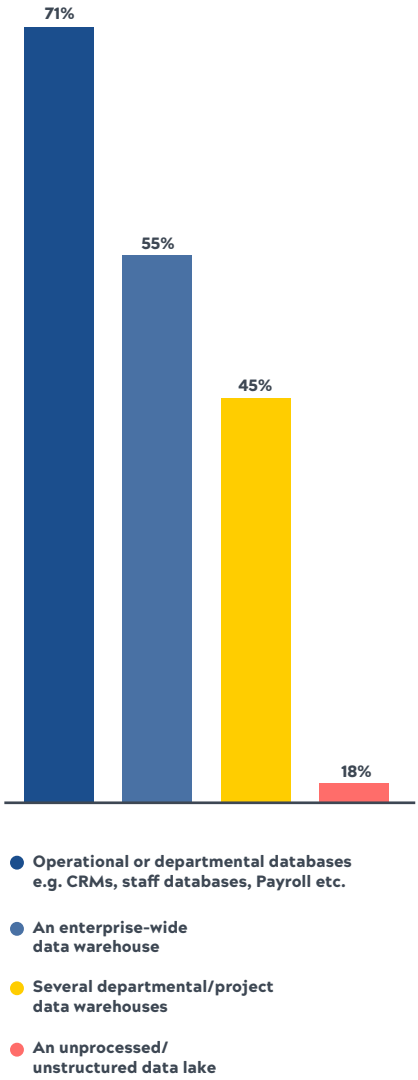

Figure 8: Which of the following does your organization currently use to store or manage its data? [500]

Added to this is the vast number of applications in use within organizations - ERPs, CRMs, collaboration apps, analytic tools, and so on - spanning cloud, on-premises, and hybrid environments. On average, organizations have 115 distinct applications or data sources (Fig.9), and only half of these (51%) are integrated so they can work seamlessly together. This means an enormous number of sources are siloed and disconnected, leaving critical data on the table or presenting an incomplete picture of the business.

Distinct applications or data sources

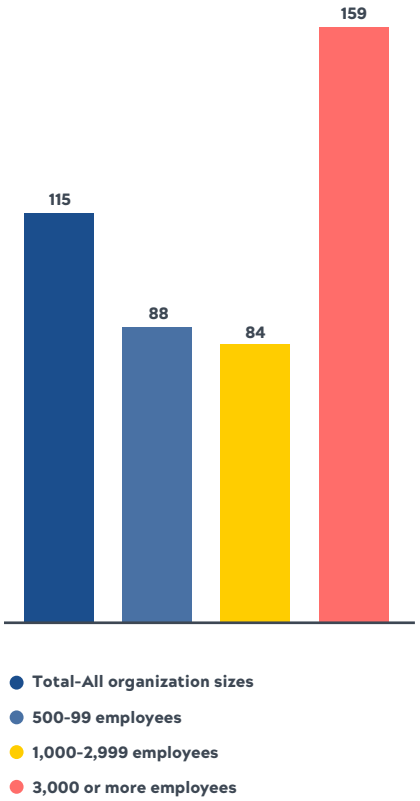

Figure 9: Average number of distinct applications or data sources which exist across organizations globally [500] Split by respondent organization size

Siloed data sources create a hurdle at the first step of data management and data warehousing. If tools aren't able to access all the data in the organization, then tasks such as data transformation and analysis are ineffective. Where the data is being located and integrated manually, employee time is wasted and the risk of human error increases.

In fact, of those who haven't integrated all of their data sources there are concerns that these data silos are preventing them from achieving their data-driven ambitions (Fig.10). Indeed, 89% believe data silos are holding them back.

Concern around data silos

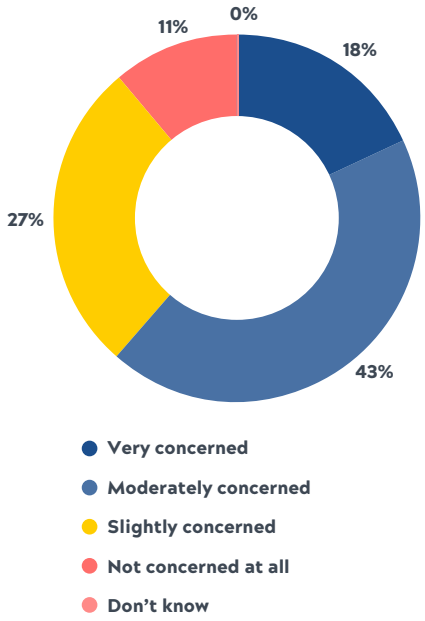

Figure 10: How concerned are you that data silos and disconnected apps/systems are holding your organization back? [466] Respondents from organizations where applications and data sources aren't 100% integrated

For IT and business teams, ensuring that their tools are able to access all the data in their organization is essential to enabling good data management practices. The importance of fast, ready access to all data cannot be overstated.

Challenges loading the data warehouse

But data silos are only part of the problem ITDMs are facing. For 49%, legacy technology is restricting data movement and the loading of data into a data warehouse, while 44% struggle with complex data types and formats (Fig.11).

Challenges of loading data

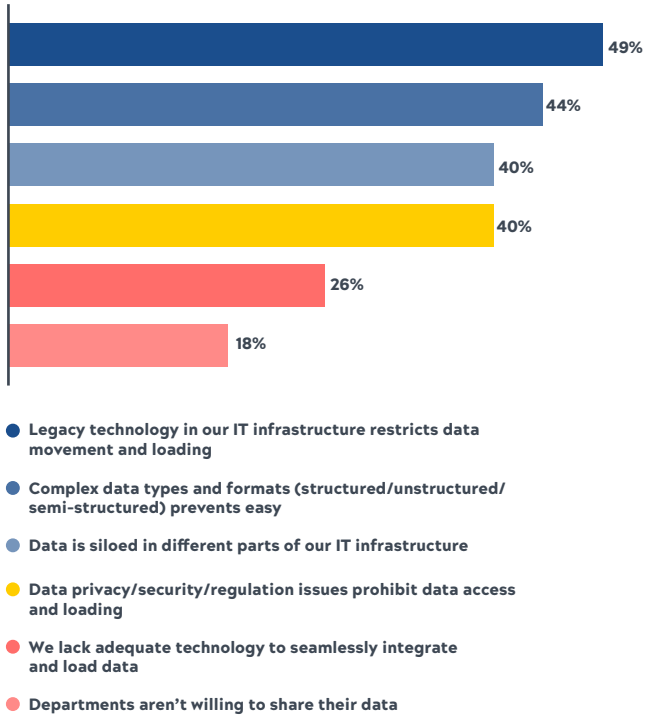

Figure 11: Analysis of factors inhibiting organization's ability to load data into data storage/management solution(s), e.g. data warehouses, data lakes, operational databases. [494] Respondents who know which data storage/management solution is in use in their organization, omitting some answers

Currently organizations are using an average of four different tools to integrate and move data in and out of data storage and warehouse solutions. Using different data integration and movement tools introduces risk, duplication, and error. If data integration and migration tasks aren't completed effectively and in a consistent manner, the process of ensuring that the data required for analytics is accessible becomes even more time-consuming and resource-draining. Further, the likelihood of data being ready for analytics when it is needed is decreased.

Time-consuming data prep

ITDMs report that, on average, 48% of the data that is migrated into their data management and warehousing solutions requires cleaning before it can be useful (Fig.12). Ultimately, this means that IT teams are losing 4 hours on average per week per employee (Fig.13) trying to solve issues such as missing data, duplicate data, or reformatting and cleaning data, negatively impacting user productivity and impeding fast data analysis.

Data cleaning required

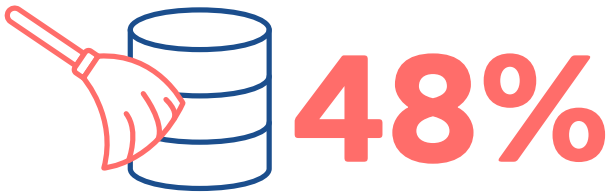

Figure 12: Analysis of data that is migrated into organizations' data storage/management solutions and requires cleaning before it is useful, respondents who know which data storage/management solution is in use in their organization [494]

Time lost

Figure 13: Analysis of hours lost on average per week, per employee, across the IT department/data team resolving issues such as missing data, reformatting and cleaning data

As long as these issues persist, the data that is held in data warehouses is unlikely to be of good enough quality to conduct robust analysis, and organizational decision making will be flawed.

Given all of this, it's not surprising that 93% of ITDMs believe improvement is needed in how they collect, manage, store, and analyze data (Fig.14).

Improvements in data management

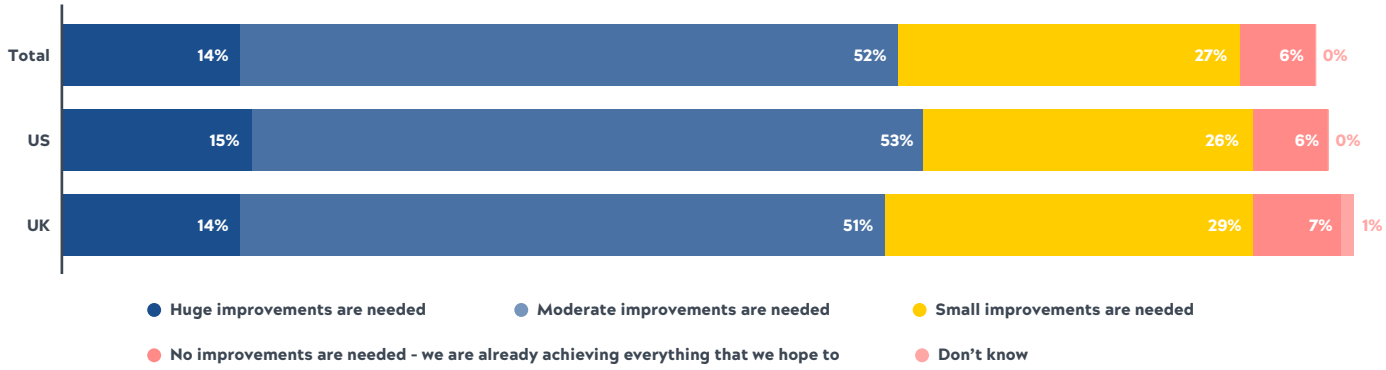

Figure 14: To what extent does your organization need to improve its approaches to collecting, managing and analyzing data? [500] Split by respondent country

The road to better data warehousing

Despite market, organizational, and technical challenges, businesses are continuing to increase their investments in data projects. So how can ITDMs look to allocate their budget to better address their data management problems and achieve their data-driven goals?

Automating data management processes

One key area which is ripe for improvement in the vast majority of organizations is the level of automation surrounding data management. ITDMs noted that 42% of data management and warehousing processes that could be automated are currently being done manually (Fig.15). For every process that is being done manually but could be automated, valuable time, resources, and money are being squandered, or worse, business opportunities are going unrealized.

Automation potential

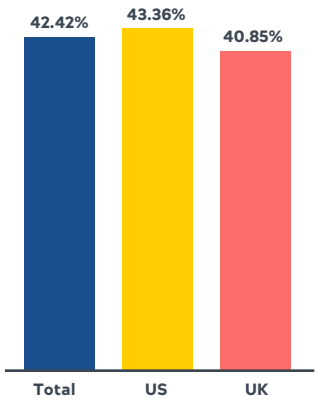

Figure 15: Average percentage of processes surrounding respondents' organizations' data management/storage solution(s) that could be automated, that are currently done manually [494] Respondents who know which data storage/management solution is in use in their organization, split by country

By automating and accelerating manual, time-consuming data management processes - including the integration of disparate applications and data sources, and the movement of quality data into the data warehouse - companies are able to speed up innovation and business value while saving considerable time and money.

Automation can accelerate all stages of data management and data warehousing, including data collection, integration, preparation, storage, sharing, and analysis. For example, automation can speed up the identification and integration of new data sources, or migration of data from legacy systems. It can automatically detect duplicate, erroneous, or missing data, or identify structures or formats that don't match the data model. And it can accelerate the loading of data into the warehouse, speeding up the data-to-decisions process.

The benefits of data management automation are many, including:

- Improved data quality and consistency, and reduced errors
- Better data monitoring, governance, compliance, and control
- Increased user productivity, and optimization of skillsets
- Increased efficiency, agility, and faster time to value
- Reduced costs and higher ROI
- Faster, better data-driven insights

By automating data management processes, IT and business teams have more time to focus on turning data-driven insights into strategic actions that will move the business forward.

7 steps to data warehousing success

Automating as much of the data management process as possible is key. What else should you consider before jump-starting your next data warehousing initiative? Here are 7 steps to get you started:

snaplogic

Step 1: Establish a data culture
An important first step is establishing, or reinforcing, a data culture throughout your organization. A data culture is, simply, when everyone in the company is switched on to the power of data. They use it, understand its potential and its limits, and know that it enables progress and accelerates success.

A strong data culture starts at the top, with the CEO and leadership team. For a data culture to thrive, data must be ingrained in the company DNA, applied in every corner of the organization, with every employee empowered by and encouraged to act on data. Good leadership, talent, training, and support are the biggest drivers for shifting culture.

Step 2: Define your data objectives
Before undertaking a new data warehousing initiative, be clear on your goals and desired outcomes. Is your aim to improve the customer experience? Or drive product innovation or new business development? Or to improve operational efficiency and reduce costs? Something else, or all of the above? Understanding your goals will help you define the right strategy, assemble the right team, secure the right tools, and build out the right processes.

Step 3: Assess your existing technology toolbox
Before setting up a new data warehouse, review what data management and warehousing systems you already have in place across the organization - including on-premises and in the cloud - and identify what's working and what isn't. Do you have the right tools in place for data acquisition, integration, preparation, warehousing, analytics, and so on? Taking an inventory of what you have will shed light on where you can leverage previous tech investments, improve on existing data practices, or bring on new tools and capabilities that are fit-for-purpose.

Step 4: Define your data sources
Audit the many different applications and data sources you have in use in your organization across all departments and regions. This will help you understand who is using what, how frequently the different systems are being used, which systems are integrated and which are siloed, the variety and volume of data you will need to migrate, etc. As part of this, take time to also look at the differing data formats and structures, as this may impact how you choose to migrate it into a data warehouse and how you transform it for use in analytics.

Step 5: Integrate, automate, and load the warehouse
One of the most important steps in any journey to better data warehousing is extracting, transforming, and loading (ETL), or extracting, loading, and transforming (ELT) data from the various applications and data sources into the chosen data warehouse. Choose a modern integration solution that can be used to integrate both applications and data, whether in the cloud, on-premises, or in hybrid environments; employs an intuitive, self-service user interface that can be embraced by users of all skill levels; and is secure, scalable, and enterprise-grade such that it can support you as your business grows.

Importantly, also choose one that embeds powerful, breakthrough AI and machine learning capabilities to automate integrations and ETL/ELT, and accelerate productivity and time to value. The more of the integration and data loading process you can automate, the more your teams can stay focused on actually using the data to drive decisions and actions.

Step 6: Report and analyze
Democratize the power of your data by putting it into a visual, easy-to-use format - including charts, graphs, maps, dashboards, and more - that helps business teams self-serve data, understand trends and patterns, make informed decisions, and take fast insight-driven action. The more people in your organization who are using data every day, the faster you become a data-driven enterprise.

Step 7: Monitor, measure, iterate
Once you've operationalized your data warehouse and analytics, use your cloud- and AI-enabled platforms to monitor, measure, and iterate. Continuous learning and improvement is key to making sure your data management practices and tools are running optimally, and that you're getting maximum value from your data investments.

Next steps

For more information about how integration and automation can support your data management and data warehousing strategy, visit **snaplogic.com**.

Research scope and methodology

The research was conducted by Vanson Bourne, an independent specialist in market research for the technology sector in May 2020.

500 interviews for this research were conducted with IT Decision Makers. To qualify for the research, respondents' organizations could be from any sector and their organization had to have at least 500 employees.

Respondents came from the US and UK:

- US - 300 interviews
- UK - 200 interviews

Respondents were interviewed using an online methodology and a robust multi-level screening process was used to ensure only appropriate respondents participated in the project.

VansonBourne

SnapLogic powers the automated enterprise. The company's self-service, AI-powered integration platform helps organizations connect applications and data sources, automate common workflows and business processes, and deliver exceptional experiences for customers, partners, and employees. Thousands of enterprisesaround the world rely on the SnapLogic platform to integrate, automate, and transform their business. Learn more at **snaplogic.com**.

Vanson Bourne is an independent specialist in market research for the technology sector. Their reputation for robust and credible research-based analysis is founded upon rigorous research principles and their ability to seek the opinions of senior decision makers across technical and business functions, in all business sectors and all major markets. For more information, visit **vansonbourne.com**

 | info@snaplogic.com | snaplogic.com

snapLogic

WP202008-L