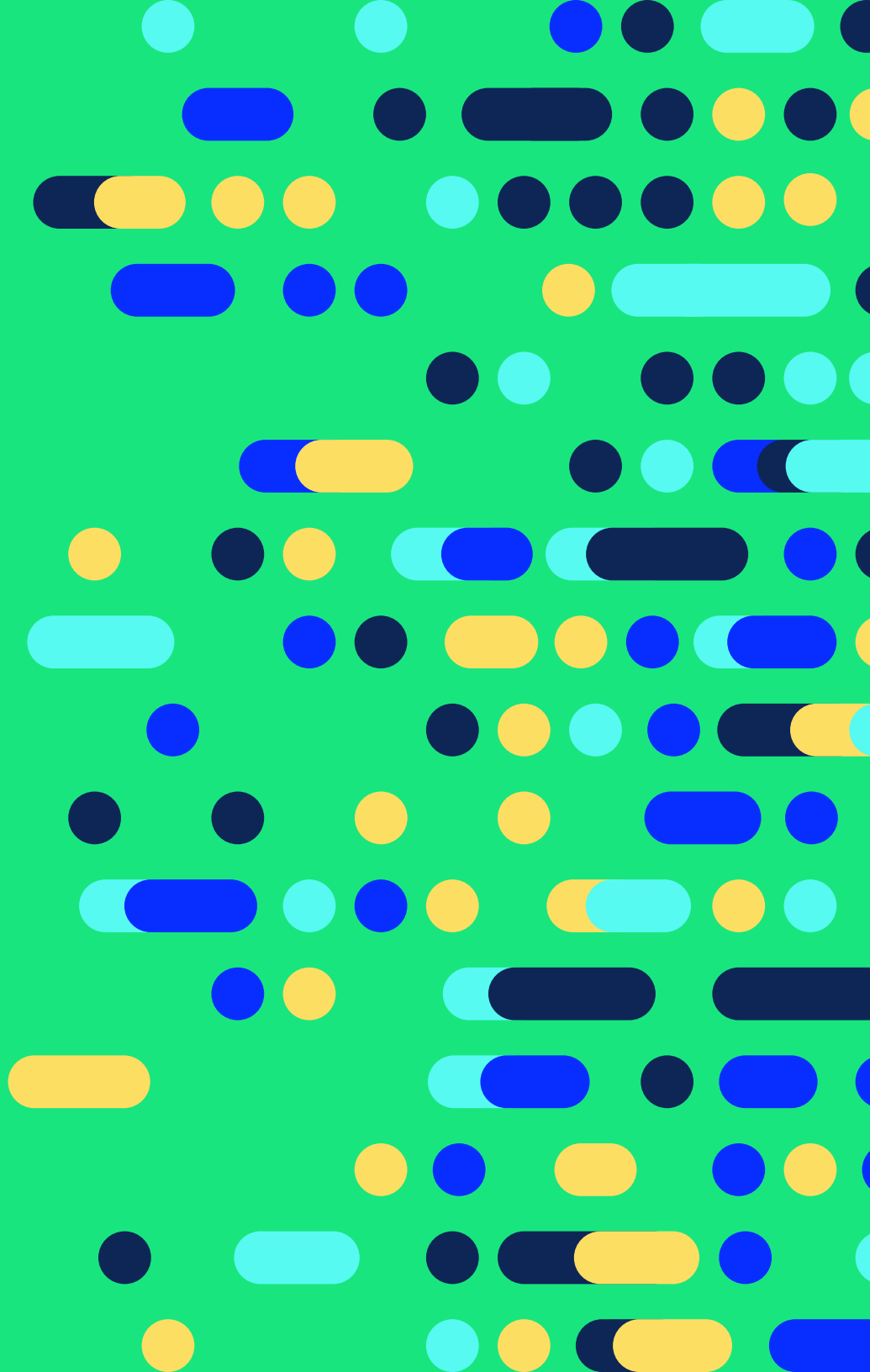


# Data Productivity:

A Survey of  
Data Experts



# Table of Contents

pg**02** **FOREWORD**

Andrew Brust

pg**05** **DATA PRODUCTIVITY PROBLEMS**

pg**07** **CAPACITY AND EFFICIENCY**

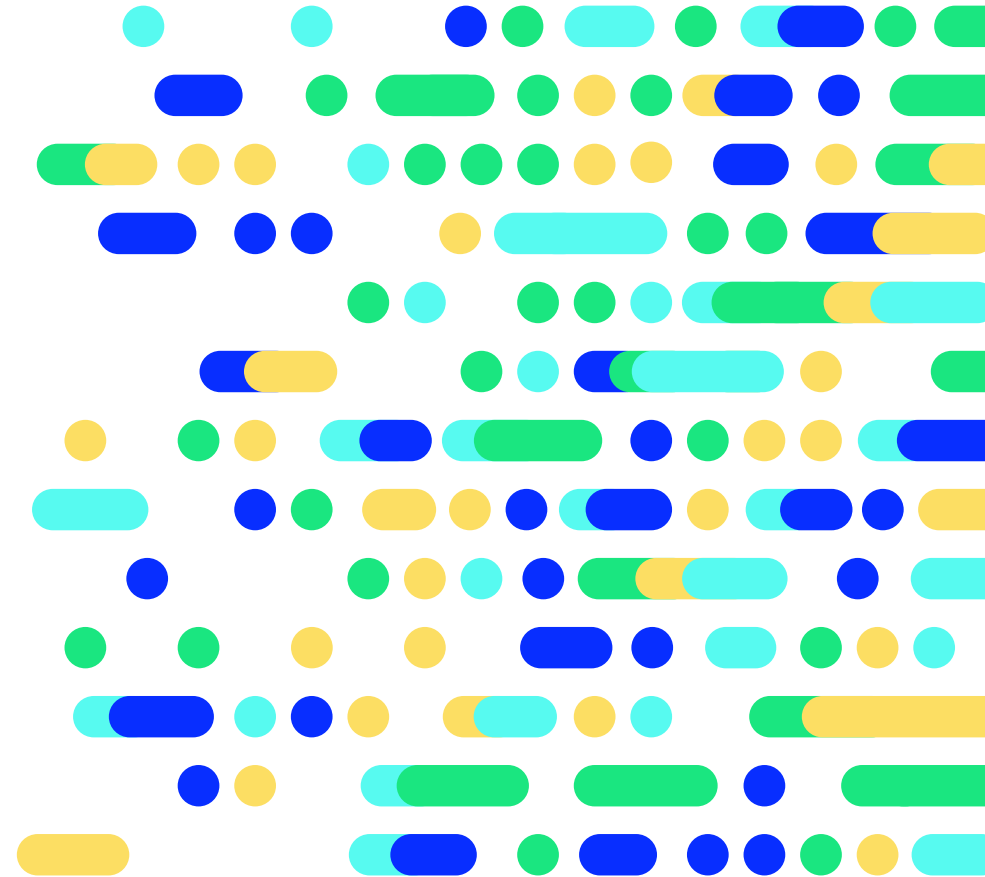
pg**09** **DATA SOURCES AND TOOLS**

pg**11** **BURNOUT IS REAL**

pg**12** **THE CLOUD AND LOW-CODE**

pg**14** **FINAL THOUGHTS**

pg**16** **METHODOLOGY**



# Foreword by Andrew Brust

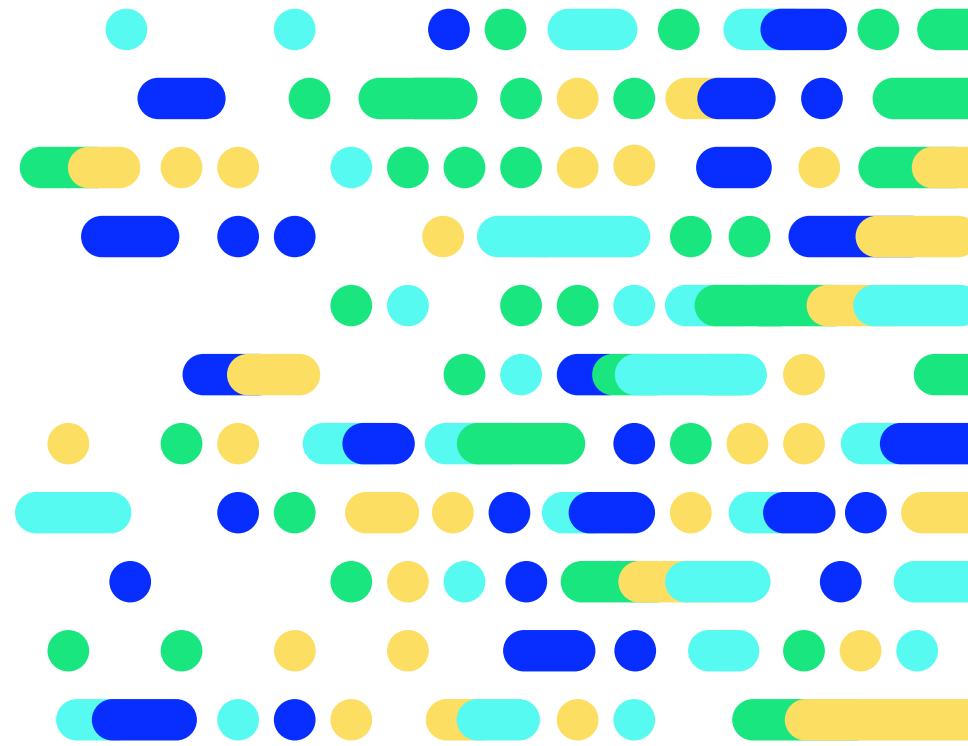
Data is everywhere today, and is growing in multiple dimensions. Not only is the aggregate volume of data expanding, but the number and variety of data sources is rapidly increasing, at an accelerating pace. This makes the need to integrate data both more urgent, and more challenging, than ever before.

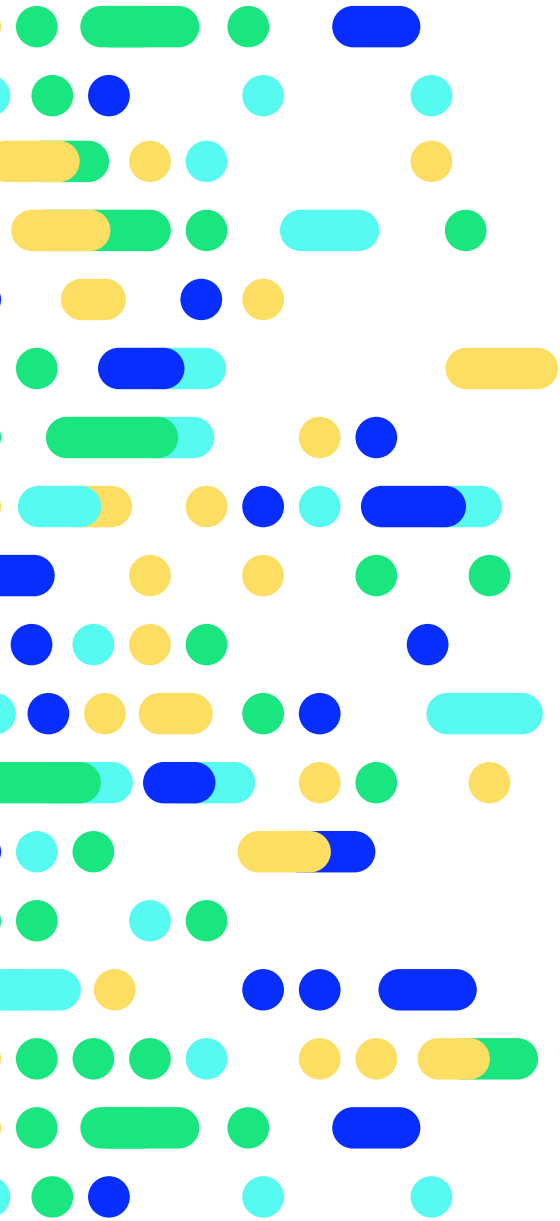
## NOT ACADEMIC

**Data integration is the key enabler to making data-driven decision making and analytics-driven business a reality, rather than mere rhetoric.** When data is integrated, discoverable and useful, it encourages people in the business to examine it, explore it, use it and master it. But when an organization's data corpus consists of little more than a bunch of files in storage or tables in a collection of databases, the idea of using it in business causes trepidation, dread and, ultimately, avoidance in the people charged with doing so.

This is a challenge that has to be solved and, if we're honest, most organizations aren't even close to doing so. This causes real-world problems, including **weaker competitive prowess, reduced productivity and lost revenue.** Opportunities shrink. Decision making becomes arduous and imprecise. And the notion of running an agile, efficient organization that controls and leverages its data becomes elusive.

This state of affairs also puts data teams at a great disadvantage. It overwhelms them with work. It buries them in the minutiae of data "plumbing," instead of the high-value and engaging work of data exploration and engineering. And because data workloads are only growing, it can cause stress that impacts morale and dilutes the very productivity these teams need to take on data sprawl.





## AGE-OLD PROBLEM, NO LONGER “BRUTE FORCEABLE”

The data integration problem isn't new. Integrating data from multiple sources into a single data warehouse for analytics was a challenge 20 years ago, too. But at that time, even large enterprises only needed to deal with an array of data sources in the single- or low double-digits. Ad hoc, hand-coded solutions, even if inelegant, could work. But in today's environment of multiple databases, files, feeds, SaaS applications and APIs, all of which serve as sources for data, the patchwork approach is simply unsustainable.

**We have to use smarts, efficiency and strategy to help us take on the data integration challenge.** And because data volumes, while formidable, aren't the only obstacle, you can't just throw hardware at this problem. Instead, you have to be sharp, proactive and forward-looking.

### WHAT TO DO

Integrating data effectively across sources requires a few things. First, you need to **give up on the idea of moving all the data into one physical location.** There are simply too many sources to do this and, even if you can do it now, it's just a matter of time before you can't. Even if the effort required to build data movement pipelines were zero, there would still be the issue of managing all of them, and dealing with the inherent brittleness associated with each one.

Second, it's important to **standardize.** The modern data stack is largely a disaggregated one, in response to the monolithic nature of the mega-vendor stacks that preceded it. While that's good for choice and flexibility, it has led to a proliferation of tools within customer organizations that provide overlapping or identical functionality. In the case of data integration, that means different platforms for different pipelines, which only increases complexity and risk. Therefore, standardizing on a single data integration platform, with a breadth of capabilities, is key.

Third, data teams must **focus on real data engineering and data transformation**, rather than data movement or replication. Shaping, conforming and taxonomizing the data should be the data team's focus, so the metrics and categorical groupings most pertinent to the business are immediately available and discoverable. That's hard and rewarding work and, although it can be very technical, it's also visionary. Inspired data transformation requires a contextual understanding of the business, its goals and potential opportunities, so that corporate data enables the analysis necessary to achieve and take advantage of them.

Sometimes the work needn't even be that technical. People who understand the business well and have passion around working with its data – even if they lack strong technology skills – can be enabled to build data transformation pipelines. **Low-code tools** have made this possible for building apps and now they're making it feasible for business users to transform and integrate data. Even skilled coders may find that low-code tools increase their productivity for more conventional, routine data transformation work, leaving them the time and flexibility needed to hand-code more advanced data transformations.

## NOW, READ THE PROOF

If all of this sounds like platitude, or just a bit abstract, read on. The explicit, quantified results from Matillion's survey will bear out the challenge and discuss its resolution in very concrete terms. Yes, hype around data has been around for quite a while. But that doesn't mean the proclamations around the power of data are wrong. What it does mean is that we must transcend the logistical difficulties in integrating data and making it ready for analysis.



**Technology like Matillion's, combined with the right priorities and investments from customers, can move us from hype to accomplishment and testimonial proof of data's utility, value and power.**

— **Andrew Brust, Founder and CEO, Blue Badge Insights**

# Data Productivity Problems

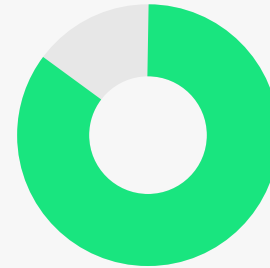
Over the past few decades, major advancements in computing and technology have created an unprecedented amount of data for businesses across every industry. But the volume of that data puts a strain on the teams tasked with integrating and delivering it.

This creates a unique challenge: Data plays an important part in helping make business decisions, measure success, keep operations firing on all cylinders and plan for the future, but data teams lack the time, cohesive tools and efficiencies to mine the data for meaningful information in a timely fashion. And with data products becoming more sophisticated now than ever, there's additional pressure for experts to deliver that data in a specific, transformed way.

Basically, these **large data sets put pressure on data teams to perform more work** — and in less time.

Matillion surveyed 900 data practitioners and decision-makers in the United States and the United Kingdom to understand where these data experts see that influx of needed data making a significant difference in their workloads. The results were significant, and draw the same conclusion...

## DATA TEAMS ARE OVEREXTENDING THEMSELVES TO MEET BUSINESS DEMANDS




**84%**  
of respondents reported  
the volume of their  
data team's workload  
exceeds capacity.



**90%**  
said their  
workload has  
increased  
in the last year.

But there are solutions for these problems and methods to mitigate them. In fact, a large percentage of participants reported two ways that would help with their workloads: 45% noted improved data access would help with their day-to-day work, and 43% suggested improved data integration processes would make a difference.

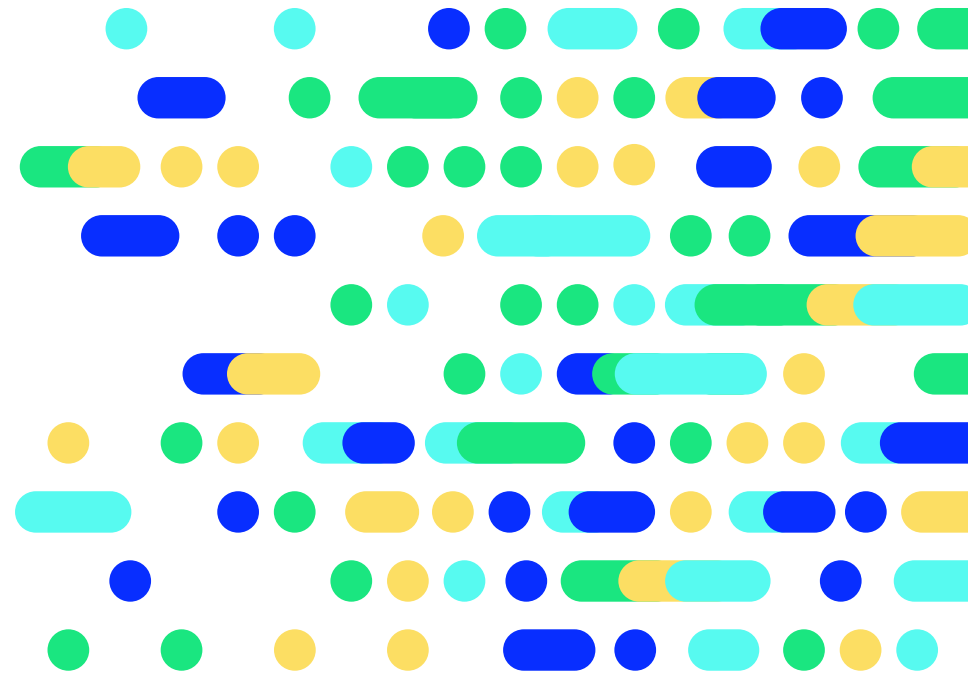
To keep up with these demands, companies have increased their spending on data movement pipelines as data products have become more complex. These movement pipelines don't necessarily add value; they bring data from one place to another alone as opposed to transforming or integrating that data. Paying for this only means companies are wasting resources on what should be a simple process. According to Wakefield Research, an average data team member spends 44% of their time maintaining data pipelines — at a cost of \$520,000 per year<sup>1</sup>.



**Companies must understand that the way they transform and integrate data costs them — bad, expensive, slow, fragmented data pipelines and a lack of collaboration within their data teams drain resources across the board.**

In fact, the IDC Global Chief Data Officer (CDO) Engagement Survey of 2021 showed that data fragmentation forces 37% of decision-makers to spend more time handling data complexity than they do driving true data transformation.<sup>2</sup>

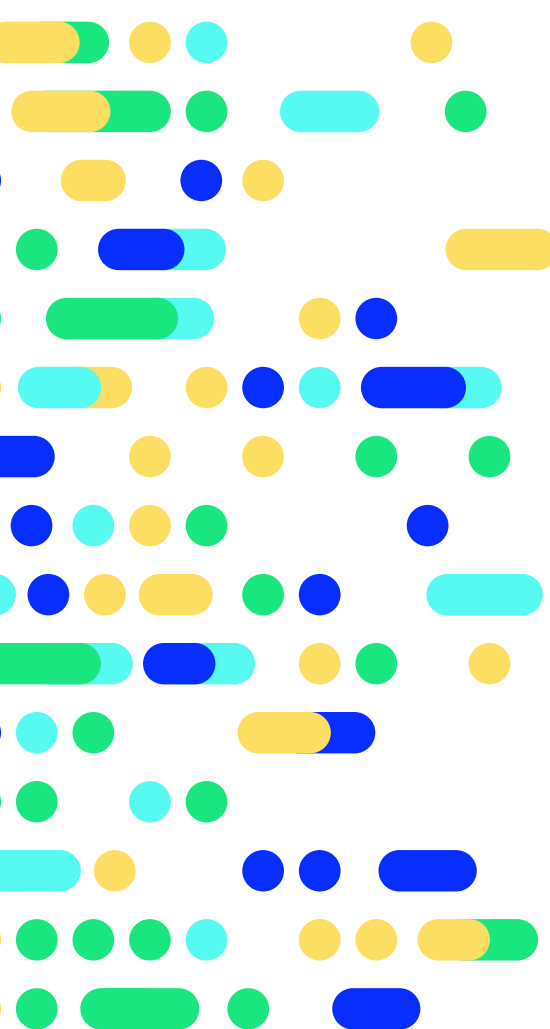
Greater insight into the impact of fragmented data and overworked data teams will only help companies figure out the right solutions.



<sup>1</sup> Fivetran, New report: Enterprises should avoid DIY pipelines

<sup>2</sup> Informatica, How Data Leadership Drives Digital Leadership

# Capacity and Efficiency



Matillion's survey found that bad data pipelines and poor, inefficient methods of data integration and transformation impact the teams who do this work. Data teams not only have to finish strategic projects that ladder up to a company's overarching goals, but have to balance that work with the more tedious routine of integrating and preparing data.

For nearly 40% of respondents, it takes between a day to a week to compile this data, and 34% reported this transformation and integration process can take between 3 to 5 hours of their day to perform. Even for teams whose processes take less time, it's still a significant lift.

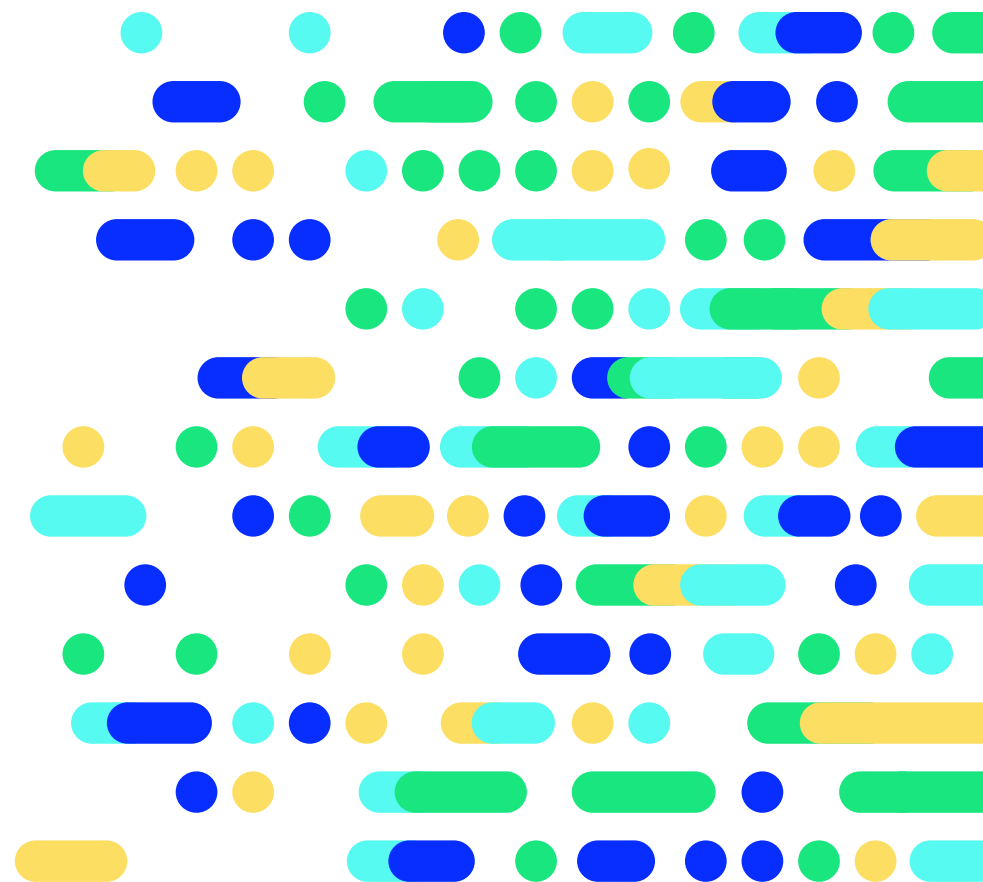


**Putting that into context, if a team member like those who participated in the survey spends 5 hours on data transformation, that's more than half of the average 8-hour workday — hours they could be putting toward more strategic projects.**



The problem, though, isn't necessarily in the amount of time it takes to do this work; it's that these pipelines are fragile and difficult to maintain. **An employee using these pipelines can't get the data they need when they need it.** Instead, they're left waiting on data engineers or developers to help them access the information they need before they can start to analyze the data: it can take developers between one to three weeks to set up a rudimentary pipeline<sup>3</sup>.

Companies have already spent massive amounts of resources, both from a team member and financial perspective on these highly technical pipelines — and instead of focusing on more complex work, data team members are having to put their focus on maintaining those pipelines. The volume of data teams are asked to sift through, and the number of tasks they're asked to perform, only put a burden on those who have the required technical expertise to access the raw, unintegrated data.



<sup>3</sup> DataChannel, Build vs. Buy Data Pipelines: A Detailed Discussion

# Data Sources and Tools



**Pop quiz:** You're asked to complete a brief data analysis assignment, but you must reference more than 50 unique resources to pull together the massive data sets before you can begin your analysis. Where do you start? How do you find a way to do this work quickly while making sure to integrate and transform the data you pull?

**Answer:** Procrastinate, swear, do your best and hope it works.

The sheer amount of tools data experts need to pull together massive amounts of data sets puts another burden on their ability to transform business-ready data.

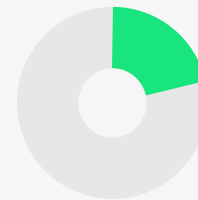
The majority of survey participants — **63%** — reported that their company uses **more than 50 different sources of data**. And according to BetterCloud's 2023 State of SaaS Ops Report, SaaS-powered companies use an **average of 186** different business-necessary apps<sup>4</sup>.

<sup>4</sup> BetterCloud, The 2023 State of SaaS Ops Report

## ADDITIONAL SURVEY RESULTS FROM MATILLION INCLUDE:



**41%**  
of respondents reported  
using 51-100 data sources



**22%**  
of respondents reported  
using 101-200 data sources

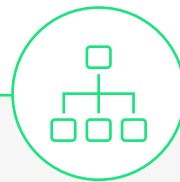


**6%**  
of respondents reported  
using 201-500 data sources

Instead of having one source for the transformation and orchestration of data integration, these experts waste their time pulling from different places, which only increases their workload.

And data experts admit that the variety of tools necessary for the task serve as a burden when they're trying to perform a high volume of work as efficiently as possible. **85% of survey participants said that using multiple data toolsets slows down their performance**, leaving them with a larger workload and less time in the day — or even week — for other projects.

Beyond the enormous time and ability it requires to use these tools, fragmented solutions create a larger impact not only on the employee assigned the task, but on the business as a whole. The variety of applications and sources can lead to vendor lock-in, reducing a team's ability to stay agile and adaptable with any newly introduced technology. They also make up a rigid, complex tech stack that becomes difficult for anyone but these specialists to navigate, and create costly data engineering practices. When a company has already spent an exorbitant amount of money and time to implement a data pipeline, it's understandable that they'd want to maintain that pipeline as much as possible.



### HOW TO SAVE YOUR RESOURCES

**Companies looking to save money and better enable their technical workers should look to reduce and simplify the number and accessibility of their data tools. Fragmented pipelines from a variety of vendors have created difficulties in data transformation that could be solved by using one simple, unified solution rather than several different ones.**

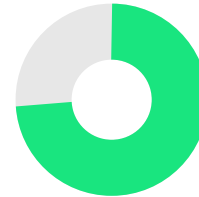
# Burnout is Real

We know that data teams are required to spend significant amounts of crucial work time integrating and moving data sets, so it makes the findings about their increased workloads admittedly unsurprising. These team members are having to put their time and focus on data transformation rather than their other important projects, which can lead to burnout and dissatisfaction.

While reports of burnout still remain manageable, it's likely that will change as businesses continue needing these critical data insights.

**More than a third of participants reported experiencing at least a little burnout, and 19% described feeling very burned out in their current role.** And it's worth noting that these numbers will only grow unless organizations make plans to solve for the problems creating this workplace fatigue.

However, it's important to note that these data experts still feel excited about the work they do and are enthusiastic about their jobs.



**74%**

**of our survey respondents described feeling very motivated about their jobs.**

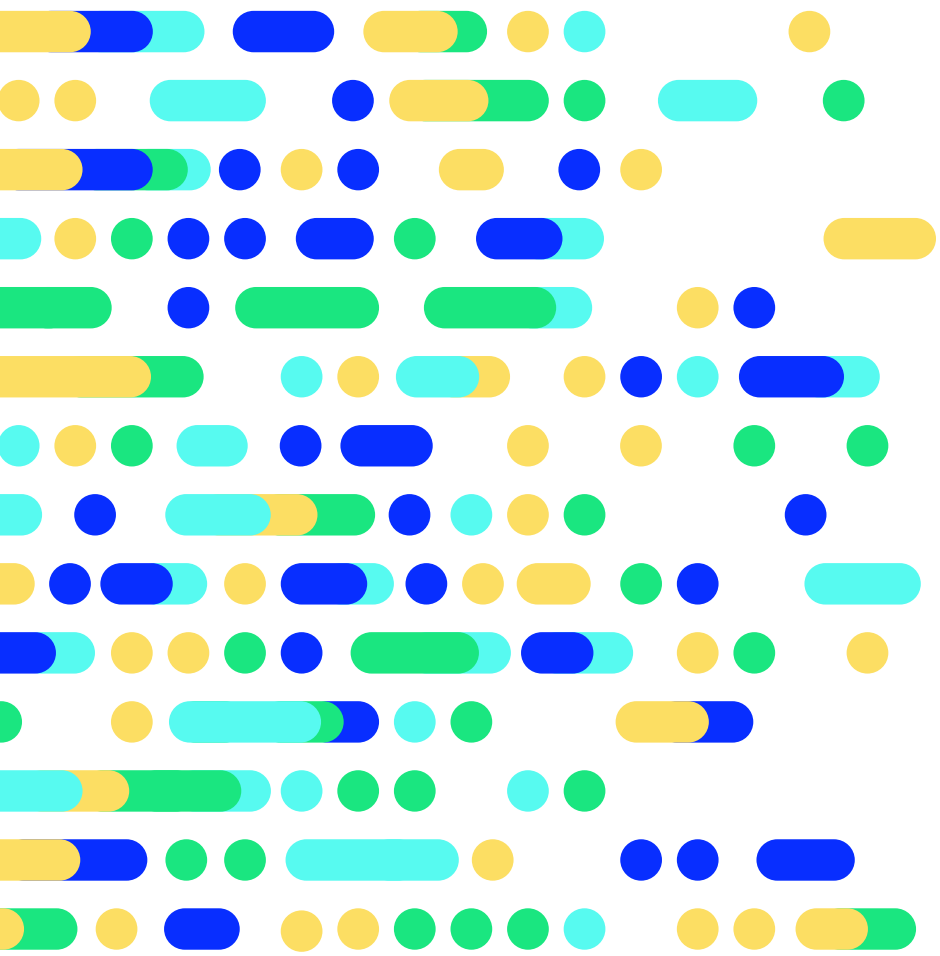
The responsibility lies on companies to relieve these employees and ensure they continue to feel driven about their contributions to the business. The more time, effort and resources leadership can put into mitigating the factors that lead to burnout, the more likely they'll be to retain top talent and help their employees perform at their very best.

To that point, burned-out data team members are more likely to leave their roles than those who feel like they have a manageable workload. And turnover on a data team runs the risk of knowledge gaps surrounding how or why things are done. All this is in addition to the already-lengthy amount of time it takes for teams to integrate and pull together data through using a large number of complicated tools.

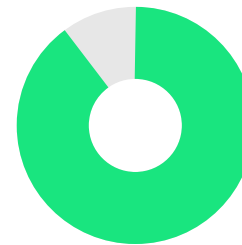
Preventing turnover, too, is a difficulty for plenty of companies; according to Gartner, **51% of organizations called talent retention one of their greatest challenges**<sup>5</sup>.

<sup>5</sup> Gartner, How to Safeguard Institutional Knowledge in the Face of the Great Resignation

# The Cloud and Low-Code



With the incredible amount of sources companies use to track and compile data, the cloud is more important now than ever.



Nearly **90%** of participants said more than half of their company's data is stored in the cloud — which is a significant amount of information.

Consider the benefits the cloud can present. The cloud in particular makes a good home for low-code platforms. Meaning: low-code and no-code can be critical components of a successful cloud solution while enabling **greater collaboration and knowledge-sharing** between teams and team members. These solutions allow quick application development that's easy for users without the technical know-how to learn and adopt. This ultimately could relieve work on data specialists as well as allow those without the coding expertise to complete necessary assignments.

**Both coders and non-coders can benefit from these solutions**, which provide coders with streamlined data integration processes and non-coders with methods of finding and transforming what they need themselves. **Sophisticated, visually guided low-code environments are instrumental in creating accurate, maintainable and self-documenting code to pull data.** This is both easier for new users to learn as well as easier to preserve, sustain and improve over time.

Low-code solutions also give data experts a way to focus less on the mundane aspects of data collection and transformation, and instead give them the ability to focus on other projects that help propel a business forward. Beyond that, these solutions are **cost-effective and agile** in comparison with many other tools specialists use to do their work today.



Another benefit these solutions provide comes with their ability to help end tribal knowledge. When a knowledge worker leaves their company, it takes an **average of 8 to 12 weeks for the organization to replace them, and another 4 to 8 weeks minimum to bring them up to speed** with the correct training<sup>6</sup>. With low-code solutions, companies no longer lose complex code when a team experiences turnover. And for coders, these solutions are much more scalable.

<sup>6</sup> Bloomfire, The Real Cost of Losing an Employee



## LOW-CODE IS A SOLUTION

Low-code and no-code provide ways to alleviate work on behalf of data teams, and it empowers those who lack the coding know-how and expertise to compile the data they may need to perform their own duties. With so many companies storing so much information in the cloud, these solutions make a natural fit for businesses, as they allow employees without technical expertise to work with, integrate and transform data for their own work purposes.

# Final Thoughts



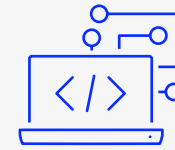
Work overload distracts data teams from their other important tasks



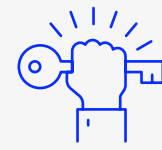
Burned-out teams don't feel compelled to stay the course at their jobs



Slow, inflexible pipelines cost money and create delays



Low-code and no-code solutions alleviate some of the pressure



Data teams need solutions that alleviate the workload

The amount of pressure data teams feel to perform quickly and efficiently despite the multitude of other projects on their plate is significant, and growing. Data experts are overworked; their assignments are often tedious and require a great amount of time, which **directly impacts performance and distracts from other strategic goals** and projects.

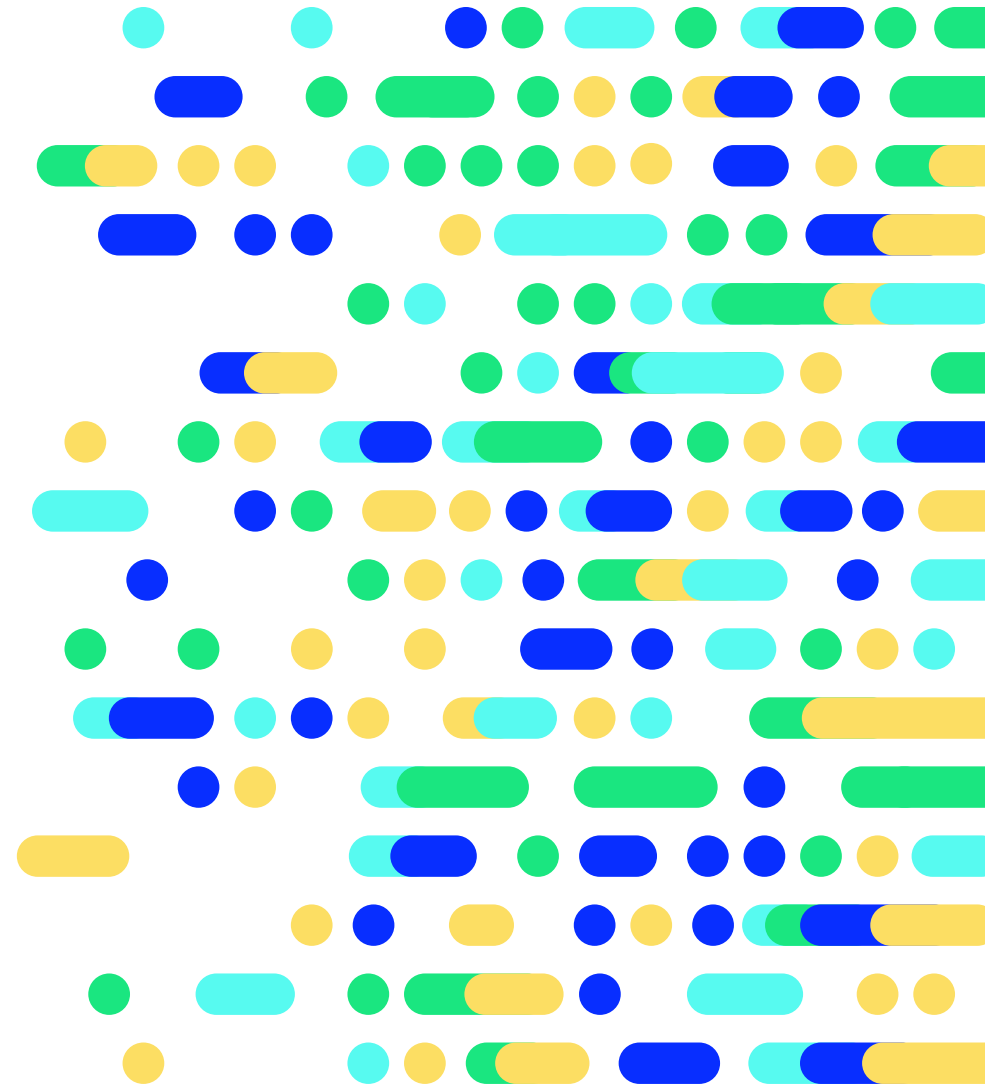
This is further magnified by **slow, inflexible pipelines that create delays** for teams working to gather business data. This, coupled with **large amounts of data sources and tools** to pull these data sets, makes their work all that much more exhausting. The use of so many fragmented apps can lead to vendor lock-in and prevent the kind of agility and flexibility needed to perform this work efficiently, and complex, rigid tech stacks lead to **costly data engineering practices**.

All of this results in a **larger workload for data teams**.

Although these employees have yet to describe intense levels of burnout, the reality is that the work will only increase as data continues to play an increasingly important role in the way businesses make decisions and execute on their goals. It might be in its early stages, but **burnout has entered the picture** — and organizations must act on this before it becomes a significant problem.

The pressure and impact of data unification and analysis on experts is something that cannot be ignored. But there are ways to improve productivity and to help mitigate the growing problems these employees experience when performing their day-to-day work, like leaning into **low-code and no-code solutions to enable data teams to work more quickly and easily while empowering employees** without that kind of coding knowledge to access the data sets they need for their own work.

By taking note of these results, and working with companies like Matillion to enable data productivity, companies can put themselves in an excellent position to help advance their organizations and support their employees to perform at their very best.





## ABOUT VANSON BOURNE

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 [www.vansonbourne.com](http://www.vansonbourne.com).

## METHODOLOGY

InkHouse commissioned independent market research agency Vanson Bourne to conduct this research. The study surveyed 300 senior decision makers, including CIOs and CTOs, and 600 users of data analytics/transformation tools in March and April 2023.

Organizations were based across the US and UK, split in the following ways: US decision maker (200), US user (400), UK decision maker (100), UK user (200). They had a minimum of 500 employees globally or a global annual revenue of more than \$250 million and represent all public and private sectors.

All interviews were conducted using a rigorous multi-level screening process to ensure that only suitable candidates were given the opportunity to participate.