

6

EBOOK

# Pitfalls to Avoid on the Road to Becoming Data-Driven



Tamr



Is your organization aspiring to become data-driven? If you're like most modern businesses, then your answer is a resounding yes! In fact, according to a recent NewVantage Partners survey<sup>1</sup>, 99% of organizations have invested in data initiatives.

But as many organizations are finding, investment alone is not enough. Despite increasing their investments, few businesses - only 24% - claim that they have created a truly data-driven organization<sup>1</sup>.

There's no question that data-driven businesses perform better. According to research by IDC<sup>2</sup>, companies that invest in data-driven behaviors see results, including 40% improved time to market, 35% increase in new customer acquisitions, and 25% increase in employee retention. Being data-driven gives these organizations a competitive edge because they have better visibility into their operations. And this visibility leads to better, more informed decisions, more effective business strategies, and fewer wasteful expenses.

While there are many reasons why organizations struggle to become data-driven, we believe there are a few common pitfalls that derail even the best of intentions. Six, to be exact.



**Only 24%**  
**of businesses have**  
**created a data-driven**  
**organization<sup>1</sup>**



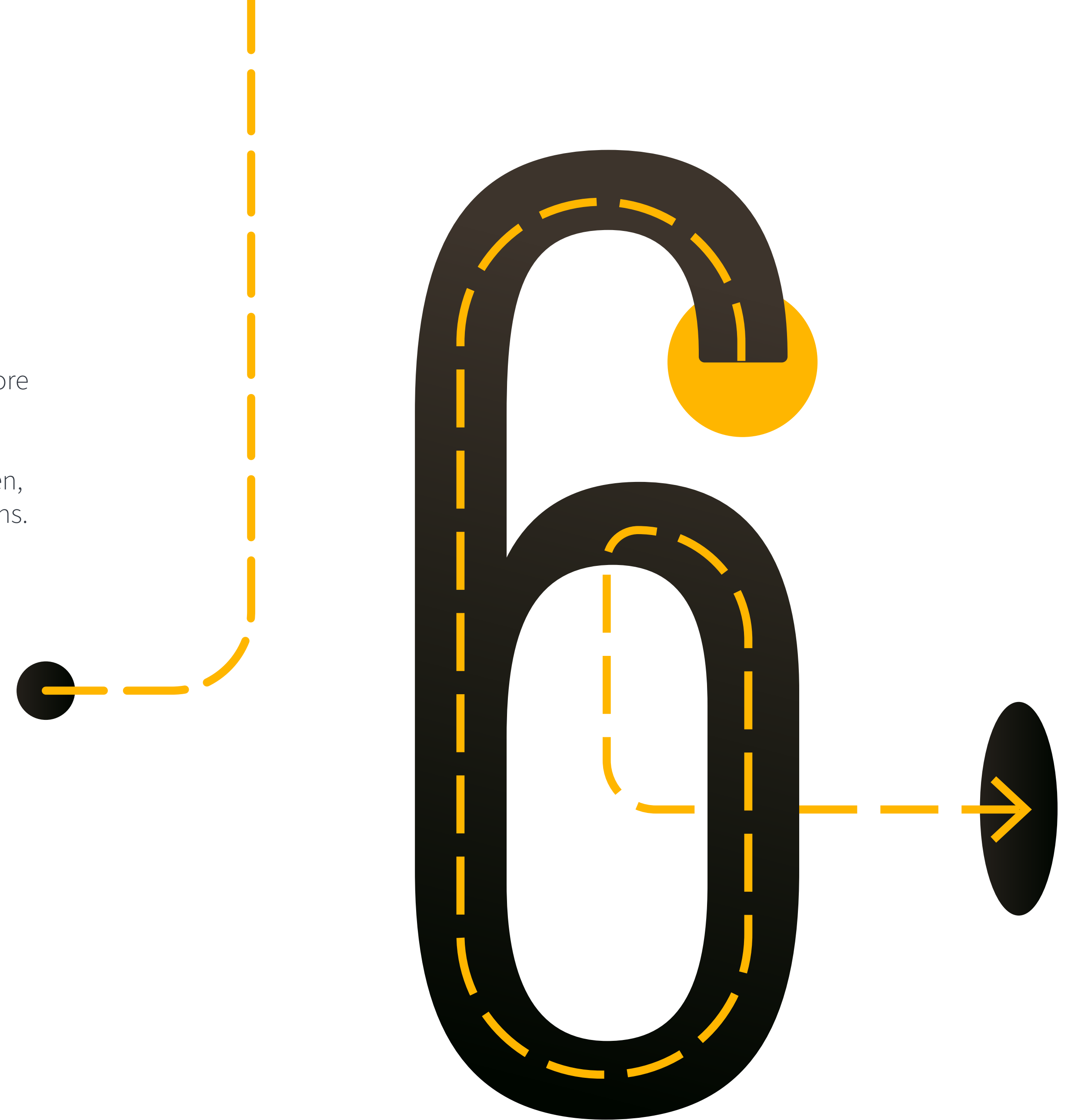
There's no question that data-driven businesses perform better. According to research by IDC2, companies that invest in data-driven behaviors see results, including 40% improved time to market, 35% increase in new customer acquisitions, and 25% increase in employee retention. Being data-driven gives these organizations a competitive edge because they have better visibility into their operations. And this visibility leads to better, more informed decisions, more effective business strategies, and fewer wasteful expenses.

While there are many reasons why organizations struggle to become data-driven, we believe there are a few common pitfalls that derail even the best of intentions. Six, to be exact.

Here they are, along with our advice on how to avoid them.

## 6 Pitfalls to Avoid on the Road to Becoming Data-Driven

- PITFALL 1 **Your solution doesn't scale**
- PITFALL 2 **Your rules are not sustainable**
- PITFALL 3 **Nobody sees the value**
- PITFALL 4 **Your architecture prevents you from real-time reading and writing**
- PITFALL 5 **Your solution lacks automation**
- PITFALL 6 **Your internal and external data are disconnected**



## → **PITFALL 1** — Your solution doesn't scale

It's no secret that data volume and variety are growing exponentially. In fact, from 2010 to 2020, the amount of data the world created, captured, copied, and consumed grew nearly 5,000%<sup>3</sup>.

Many companies are struggling to scale their data solutions to meet the demands of this ever-expanding data set. Some are using traditional master data management (MDM) to master their data. But the challenge with traditional MDM is that it requires a human-intensive process that depends heavily on manual reviews of exceptions. And that means it's not scalable as your data grows in volume and complexity.

Without a scalable solution, your business will quickly find itself unable to handle the workload that your data requires, resulting in a large and growing portion of data remaining unmastered. And you'll find that the only way to keep pace with the ever-increasing volume of data will be to invest in more human resources. And that's a costly proposal.

### **How to avoid: Invest in a cloud-native, next-gen MDM solution**

To manage this rapid, ever-expanding volume of data, companies need to take advantage of the latest cloud-based technologies to master the petabytes of data they rely on for decision making.

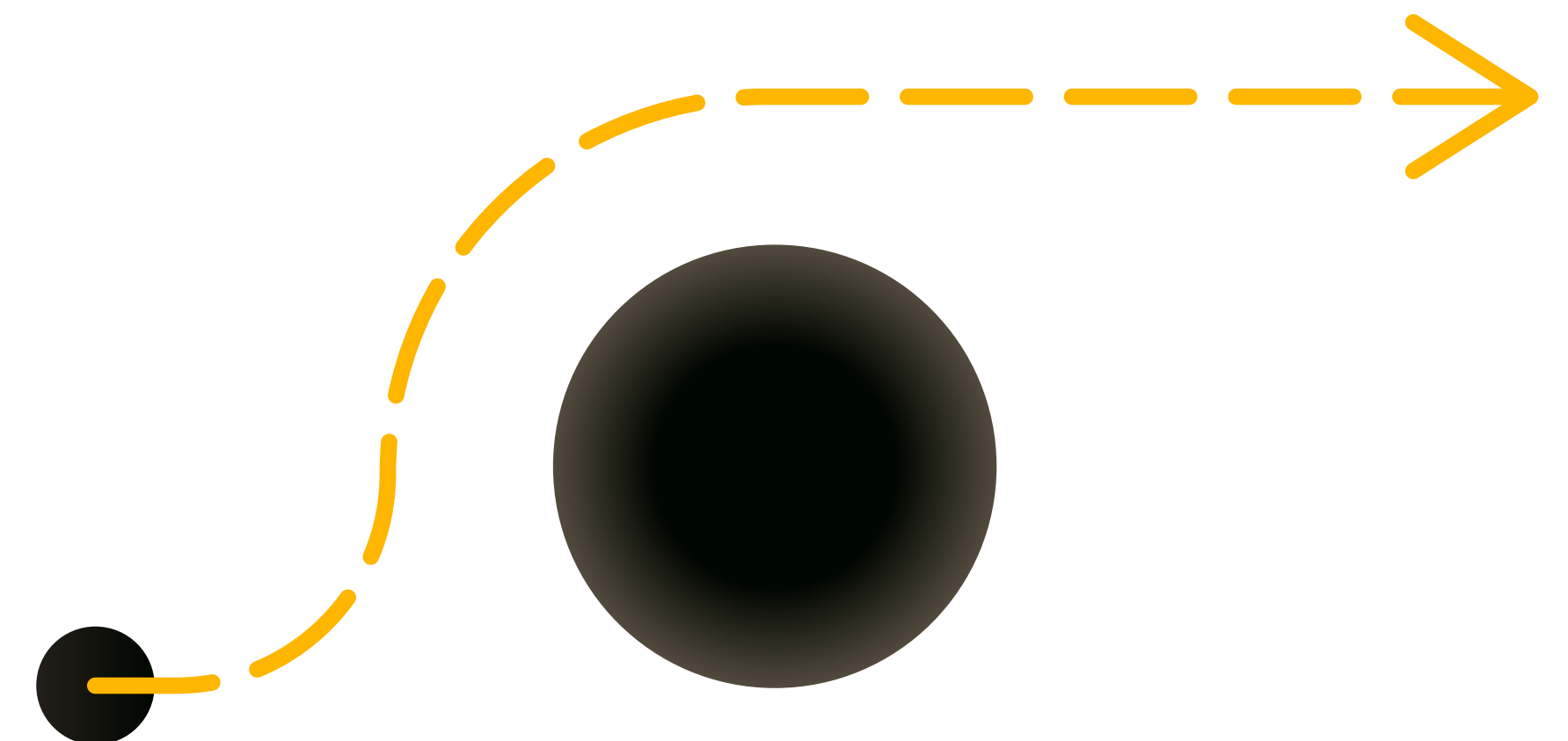
Cloud-native MDM solutions offer companies the ability to easily connect internal and external data sources so they can power analytic insights. They integrate machine learning with human feedback to break down data silos and deliver the clean, curated, and comprehensive data that data-driven organizations need to succeed.

When you invest in a cloud-native, next-gen MDM solution, you're setting your business up for success. Just look at Analog Devices.

Prior to using Tamr, Analog Devices struggled to match new leads to existing accounts, with only 36% of leads matching a corporation. Third-party match services were slow and

## Want to hear more about how Analog Devices used customer mastering for B2B sales?

 [Watch the video](#)







costly. And the organization had difficulty handling the mix of customer languages.

Analog Devices implemented Tamr's machine learning-based approach which enabled the organization to accurately match leads to accounts to drive their cross-sell and upsell account strategy. They also significantly reduced - or in some cases removed - their reliance on third-party match services through improvements in in-house enrichment capabilities. And through the use of Google translate enrichment, they tackled language challenges across locations.

Analog Devices achieved a global view of accounts that supported their cross-sell and upsell strategy. They also realized additional results, including:

- 50% increase in the match rate of sales leads to existing accounts within the first 2 weeks
- Improved lead response time from weeks to days
- Savings of hundreds of thousands of dollars on third-parties





## PITFALL 2 — Your rules are not sustainable

How many rules does your organization have? If it's over 500, then your life is about to get really hard. But if you don't believe us, then [take it from Michael Stonebraker](#).

Rule of thumb: the more data sets you're integrating, the more rules you need. And as your data sets grow incrementally, the number of rules you need grows exponentially. [Dan Waldner](#), former Director of Customer Data at Scotiabank, estimates:

8 data sets = one thousand rules

10 data sets = one million rules

20 data sets = several million rules

So if you have over 500 rules (and who doesn't?) and you're relying on humans to write these rules, then you will struggle to onboard new data sources. Why? Because **every time you add a new source, you need a human to rewrite the rules**. And as you can see, the number of rules that person needs to write is massive. When you're trying to scale, this is a surefire way of doing the complete opposite. See pitfall #1.

### How to avoid: Invest in next-gen MDM solutions that integrate machine learning with human feedback

Machine learning is critical when it comes to scalability. It does the heavy lifting, helping you to onboard more data sources faster. And it gets better with more data. In fact, the bigger the data, the more accurate the results. But it's not something you can simply bolt-on at the end to inform how humans write the rules.

Instead, machine learning must be at the foundation of any modern MDM solution. That way, humans can spend their time guiding the machine and building trust through bottom-up feedback that considers the context of how their organization is consuming the data. The machine will use this human subject matter expertise to learn over time. And by time, we mean about a week.

Societe Generale, one of the leading European financial services groups, struggled to gain a unified, up-to-date view of their €6.5 billion in global spending. Previous, rules-based solutions

didn't scale, connecting only 15 of the 100+ ERPs currently in use across the firm and providing only 40% visibility into spend worldwide. And connecting new data sources was costly and time-intensive, adding to the scalability issues.

By partnering with Tamr, Societe Generale was able to see big results in less than two months. Using human-guided machine learning, they mastered their supplier data to gain their first-ever unified view of global spending and reduced manual data preparation by 90%. They added new sources within days (not months) and reached a 70% confidence in classification, all with minimal training.

Within three months, the firm drove procurement savings by generating accurate views of the company's spend for 300+ groups. And, the IT staff was able to focus on higher-value projects, not data preparation.

**See how Societe Generale unified their spend data using Tamr's human-guided machine learning.**



**Watch a 3-minute video**

## PITFALL 3 — Nobody sees the value

There are many reasons why master data management projects fail. But the #1 reason is that most projects lack the structured framework needed to qualify and quantify data management value creation.

That's why a pitfall many project leaders fall prey to is failing to [build a business case](#). Without a business case, the team has no clearly-defined goals. Nor do they have a defined way to measure success against those goals.

Without these goals and measurement frameworks in place, you will never know whether the project was successful - or if it was not. Moreover, executive and leadership teams will fail to see the value of the hard work invested by the data teams because there is simply no way to demonstrate the benefits it provides.

### How to avoid: Create a business case that defines how you'll prove the value

Successful projects, on the other hand, have a clearly-defined business case with goals that are explicit, right from the start. They build in a way to measure the value the project provides, and they secure buy-in from leadership on how they will demonstrate success.

To ensure your project starts off on the right foot, start by looking for resources that help you estimate the total return on investment (ROI) from a next-gen MDM project. Tamr recently worked with Forrester Consulting to conduct a [Total Economic Impact™ \(TEI\) study](#). This study examines the potential ROI an organization could realize by deploying Tamr and provides a framework to assess the potential financial impact of the Tamr platform on their organization.

Using the [ROI calculator](#), you can easily calculate the potential ROI for your organization. Simply adjust the numbers to align with your business and see what kind of return you could achieve.

Then, combine these projected outcomes with Tamr's Business ROI Optimization system. This [four-step process](#) simplifies and accelerates the process of creating a detailed and defensible business case that not only defines outcomes and success metrics but also measures value realized by potential savings and/or improvements in your operational efficiencies.

Share this business case with your leadership team. Get their buy-in, and you'll be one step closer to proving the value of your next-gen MDM project.



## PITFALL 4

### Your architecture prevents you from real-time reading and writing

Healthy, next-gen data ecosystems need the ability to simultaneously process data in both batch and streaming modes. And this needs to occur not only from source to consumption, but also writing back to the source, which is often an operational system (or systems) itself.

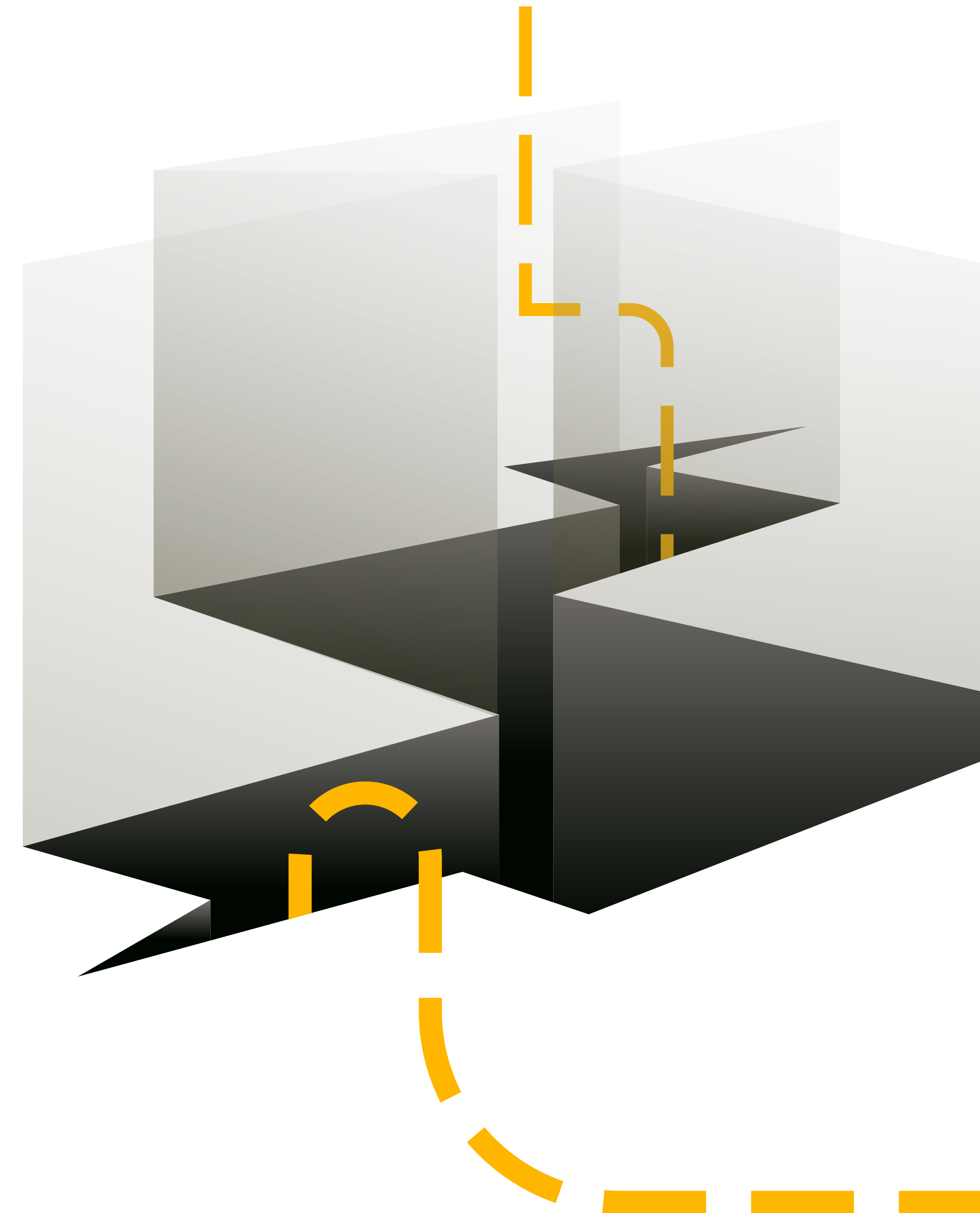
It's also important that you have a real-time engine. Not only does it enable you to provide instant results for all users and operational systems such as SAP, Salesforce, and Oracle, but having the ability to read and write at real time - or near real time - also prevents you from creating bad data in the first place.

In addition, real-time read can enable autocomplete functions to block errors at the point of entry. And real-time write through the MDM services and match index can help share good data back to the source systems.


Unfortunately, many solutions have architectures that do not allow you to have all the batch and stream and read and write capabilities. Beware of these solutions, as they will derail your efforts to write back to the source system with updated information in real-time.

**How to avoid:** Embrace solutions that allow for real-time reading and writing  
Traditional, rules-based MDM solutions hinder your ability to read and write in real-time. Instead, you need to embrace a modern MDM solution defined by human-guided machine learning. That's what one of the top three global shipping companies did, and they saw immediate impact.

Plagued by a legacy MDM solution that was slow to provide insight, a team of 60 spent five







years on a project that only mastered five out of 130 entities - and the accuracy of those mastered entities was questionable at best.

Their manual, rules-based approach failed to scale as data volume and variety continued to rise. And they found that mastering additional entities using the legacy MDM solution would not only require significantly more time, but also a substantial increase in headcount.

So they decided to partner with Tamr, and using human-guided machine learning, they generated over three million complete customer records within weeks. Their cloud-native Azure deployment drove fast and efficient processing of data at scale, generating results in weeks - not months or years.

Further, using an open API-based platform and low latency matching capabilities, they achieved seamless integration into their data architecture.

The results speak for themselves. Using Tamr, the global shipping company was able to:

- Drive millions in upsell opportunity with complete customer records
- Pursue supplier use case estimated to save millions by identifying key vendors and negotiating pricing
- Onboard new data sources in weeks (not months) with legacy rules-based MDM solution
- Drastically reduce time spent manually preparing data which allowed data experts focus on higher-value projects



## PITFALL 5 — You've burned out your data engineers

According to DataKitchen<sup>4</sup>, 97% of data engineers are feeling burnout. Why? Because they are constantly reacting to the most pressing data issue at hand. Data engineers spend much of their time fighting data fires. When customer requests, broken systems, and data errors arise, the organization sends an SOS to the data engineers and expects them to work around the clock to make things right.

Because data engineers are always reacting, they are unable to embrace continuous improvement. And this leads to frustration, burnout, and ultimately turnover. That's why burning out your data engineers is a pitfall every organization should avoid. Data engineers are a rare breed, which makes it difficult to replace them when they've simply had enough.

### How to avoid:

#### Automate as much as possible

The good news is that there is a simple way to avoid this pitfall: embrace [DataOps](#) to automate as much as possible.

Good data platforms should automate as much as possible and institutionalize knowledge so that data scientists and data engineers can spend less time processing the data and more time analyzing it. Automation also helps to solve the issue with burnout. Instead of being data firefighters, data engineers can spend time continuously improving the data and eliminating manual, time-consuming processes.

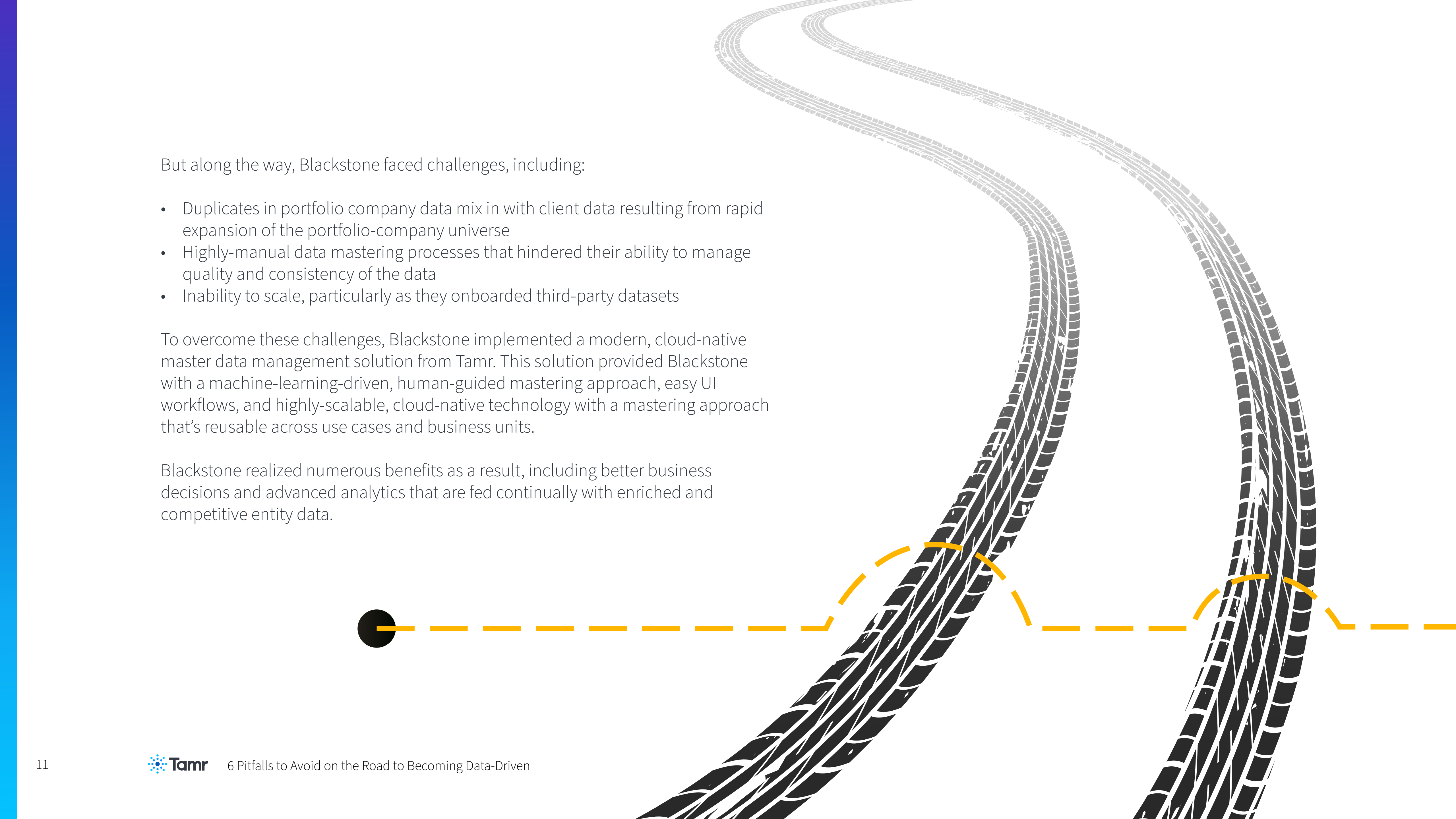
[Blackstone Group](#) operates with a cloud-first mentality and sees the value in automation when it comes to data management. And through its partnership with Tamr, Blackstone is benefiting from repeatable, scalable and reliable data mastering as an essential component of their master data management (MDM) system.

After years of amassing a huge variety of data assets across multiple systems and sources, Blackstone realized it needed to fully master its data and create golden records for all key entities.



97%  
of data  
engineers  
are feeling  
burnout<sup>4</sup>





But along the way, Blackstone faced challenges, including:

- Duplicates in portfolio company data mix in with client data resulting from rapid expansion of the portfolio-company universe
- Highly-manual data mastering processes that hindered their ability to manage quality and consistency of the data
- Inability to scale, particularly as they onboarded third-party datasets

To overcome these challenges, Blackstone implemented a modern, cloud-native master data management solution from Tamr. This solution provided Blackstone with a machine-learning-driven, human-guided mastering approach, easy UI workflows, and highly-scalable, cloud-native technology with a mastering approach that's reusable across use cases and business units.

Blackstone realized numerous benefits as a result, including better business decisions and advanced analytics that are fed continually with enriched and competitive entity data.



# PITFALL 6

## Your internal and external data are disconnected

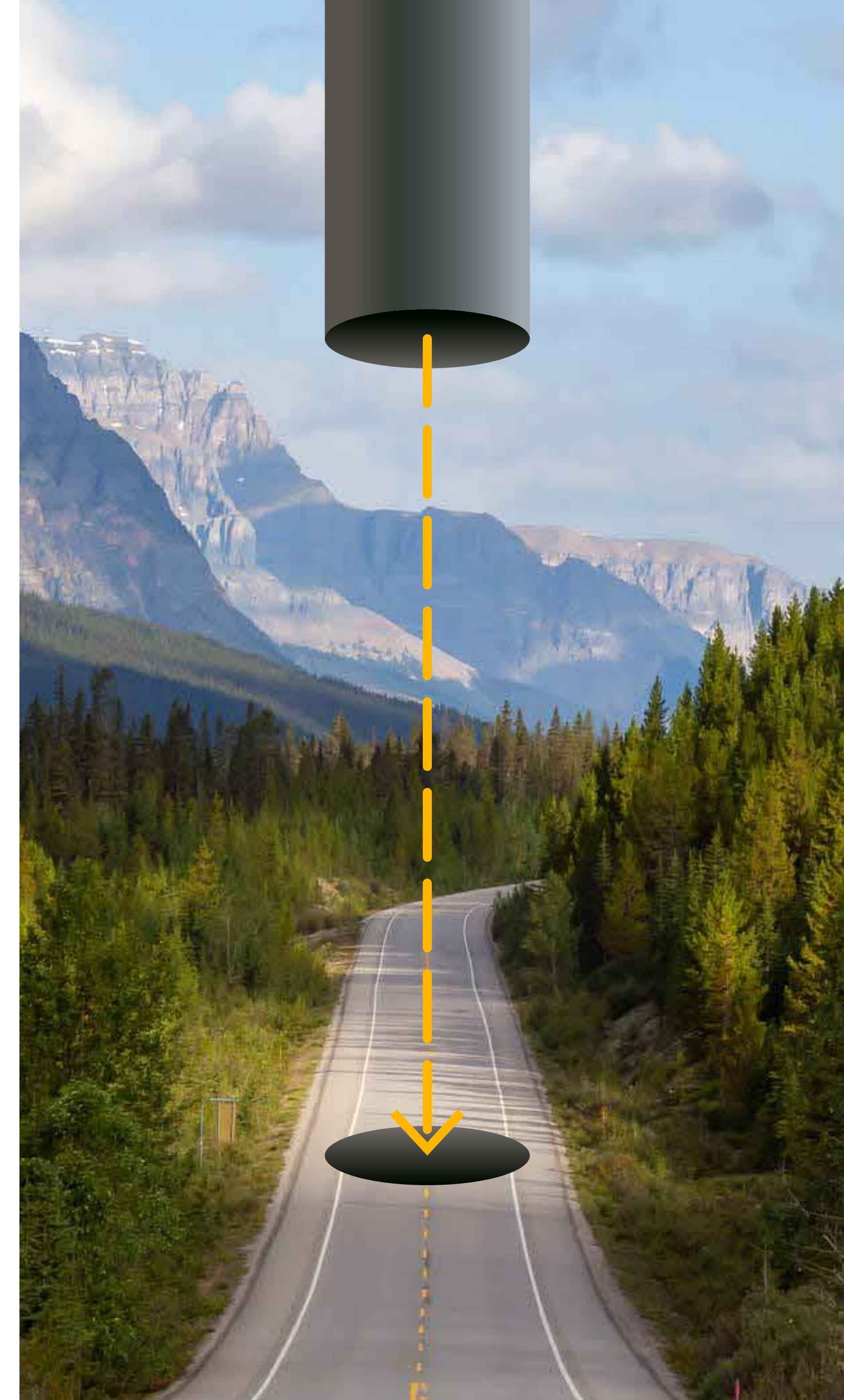
Internal operations produce a lot of data. And organizations have made huge investments and considerable strides in collecting and using this data for business decision-making.

However, for many organizations, external or third-party data remains an untapped resource. And failing to capture the full potential of linking internal data with external data remains a missed opportunity.

Take the recent COVID-19 health crisis as an example<sup>5</sup>. As the pandemic disrupted virtually every aspect of our lives - from how and where

we purchased goods and services, to the activities we engaged in, to our digital behavior - businesses found their existing consumer research, forecasts, and predictive models quickly became outdated. They struggled to understand these fast-changing patterns, and soon realized that their internal data was of little use.

Instead, they needed to augment their existing data with the wealth of external cloud data available to help them plan and respond as their business environment changed.





# HOW TO OVERCOME

## Enrich your internal data with external data

Overcoming this pitfall requires your organization to connect your wealth of internal data with valuable external data from third-parties, vendors, or public data sources relevant to your business. And by recognizing the value of this linkage, you'll gain a competitive edge in today's dynamic business environment.

Staples improved marketing and sales operations by modernizing their MDM approach and curating customer data from disparate data sources using Tamr's cloud-native data mastering solutions.

Faced with outdated and incomplete customer data, Staples found its data challenges slowed marketing initiatives and disrupted sales operations. But correcting a single customer record took months and manual coding just to adjust rules systems. This effort led to a high

total cost of ownership because they required dozens of highly-skilled resources just to maintain these rules.

In partnership with Tamr, everything changed. Staples was able to use Tamr's modern MDM solution to:

- Create complete, up-to-date customer views from millions of records across 11 Salesforce and operational data sources
- Generate unique Tamr IDs for each customer, removing the 40% duplication between Salesforce and systems
- Cluster customer views into business-ready segments by territory, industry, etc.

As a result of this effort, Staples improved marketing efficacy, boosted sales operations, and enhanced its customer experience initiatives.

Encountering MDM pitfalls on the road to becoming data-driven is inevitable. But if you're using traditional MDM, you're assuming that a steady state of data supply and demand is within sight. That with just one more integrated source, one more optimized set of integration rules, or one more multi-million dollar, multi-year contract, your organization would reach data utopia.

Don't let these obstacles derail your efforts. By putting these six principles and **modern, cloud-native data mastering from Tamr** into practice, you can easily overcome any obstacle in your path.

Learn more at [tamr.com](https://tamr.com).



- 1) New Vantage Partners, Big Data and AI Executive Survey 2021, Executive Summary of Findings
- 2) Forbes, 5 Trends that Set Data-Driven Organizations Apart
- 3) Forbes, 54 Predictions about the State of Data in 2021
- 4) DataKitchen, 10 DataOps Principles for Overcoming Data Engineer Burnout
- 5) McKinsey Digital, Harnessing the Power of External Data



Tamr is the world leader data mastering. We accelerate business outcomes for leading organizations by powering analytic insights, boosting operational efficiency, and enhancing data operations. Tamr's cloud-native solutions offer an effective alternative to traditional Master Data Management (MDM) tools, using machine learning to do the heavy lifting to consolidate, cleanse, and categorize data. Tamr is the foundation for modern DataOps at large organizations including Industry leaders like Toyota, Santander, and GSK. Backed by investors including NEA and Google Ventures, Tamr is transforming how companies get value from their data.

Learn more at **[tamr.com](https://tamr.com)**

