

# THE CATASTROPHIC COST OF BAD DATA

## AND AN IRREFUTABLE CASE FOR MULTIDIRECTIONAL SYNC

By Nick Bonfiglio, Founder and CEO of Syncari  
with Chris Gillespie, Editor in Chief of Find A Way Media

*"A must-read for the entire C-suite."*

Phil Fernandez, Marketo CEO and Founder

*"A crushing blow to the overheated  
CDP market."*

Jon Miller, Engagio CEO and Founder



# **THE CATASTROPHIC COST OF BAD DATA**

**AND AN IRREFUTABLE  
CASE FOR  
MULTIDIRECTIONAL SYNC**

By Nick Bonfiglio, Founder and CEO of Syncari  
September 2019



## THE ORIGIN

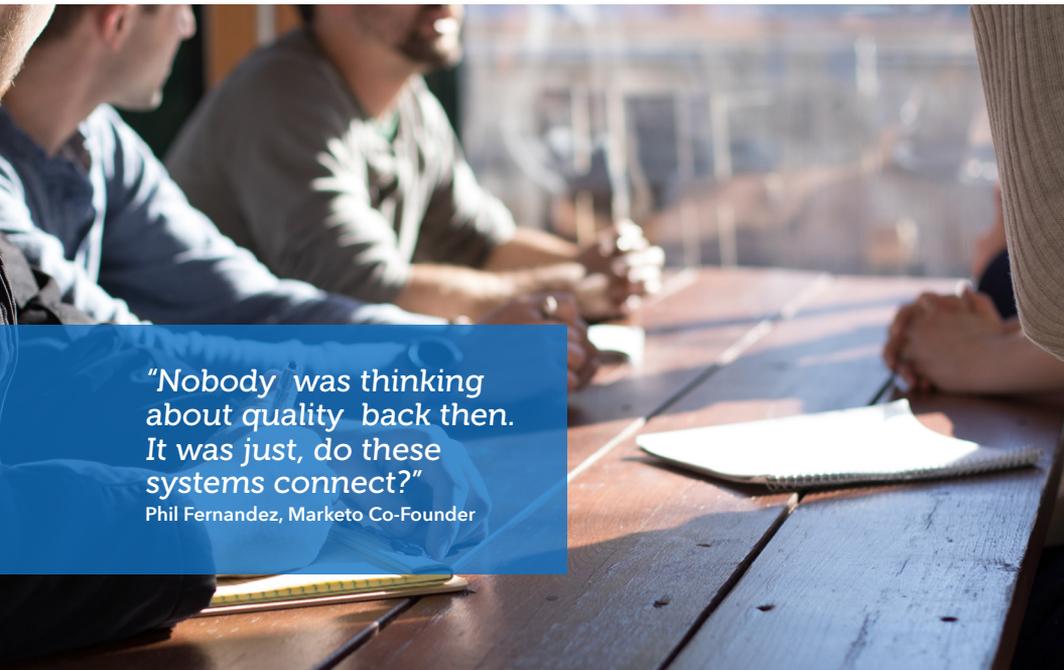
*In the fall of 2013, the entire Marketo executive product team squeezed into a conference room to discuss a list of things our users didn't want.*

We were rebuilding Marketo's sync to CRMs like Salesforce and Microsoft Dynamics, and we knew that marketers didn't want to talk to IT. They didn't want to think about which fields mapped to which or how conflicts were resolved. They just wanted to enter their credentials to start sending emails so they could achieve their true purpose: earn a seat at the revenue table. So we built a powerful, bidirectional sync for maximum convenience.

For the better part of a decade, Gartner's Magic Quadrant listed that sync as one of Marketo's top competitive advantages. It still is. Even when a competitor

was acquired by Salesforce, they couldn't build a more intuitive integration.

Fast forward to my next company, Aprinsic, where once again, we needed to build a Salesforce integration and once again, we had to start from scratch. People advised us to use cloud connectors and infrastructure tools (known as iPaaS), but those would have inconvenienced customers and in some cases, doubled the cost. Once again, out of empathy for the user, we built it ourselves and wondered, if data is so important, why doesn't anyone offer a way to make true syncing easier?



*"Nobody was thinking about quality back then. It was just, do these systems connect?"*

Phil Fernandez, Marketo Co-Founder

## THE QUESTION OF DATA QUALITY

Looking back, I realize that Marketo's integration was different because we were solving for the problem of data quality. We weren't trying to connect systems, we were trying to make sure marketers had good data. The sync was incidental to ensuring that their CRM and marketing systems agreed. That was then. Today, companies have 12 to 15 apps hooked into their CRM and the ideal of unified data seems even more distant. All those systems offer their own custom point-to-point connections which sometimes only flow one way. None have the intelligence to move data in accordance with its true purpose—to power the business.

This is more than an academic exercise. Jumbled connections are a chief cause of the \$3.1 trillion lost to bad data each year.<sup>1</sup> It can be worth 15 to 20 percent of a company's revenue to correct this problem, and no business can create a unified customer

experience without addressing it.<sup>2</sup> As Marc Benioff put it in a recent Salesforce investor call, the future is here and "every company needs an intelligent, 360-degree view of their customers. They need to personalize every customer experience. They need to predict customer behavior and anticipate customer needs." Yet how exactly does a company act on that?

*This guide will explain why so many data ecosystems are broken, the sometimes catastrophic effects, and how you can fix yours to the tune of millions in saved revenue.*

**Nick Bonfiglio**  
CEO, Syncari



## EXECUTIVE SUMMARY

Most companies have **12 to 15** core systems hooked into the CRM, and growing.

Most of those connections are **point-to-point** and prevent companies from managing data quality across the business.

Only **16%** of managers trust the data they work with and bad data costs companies **15% to 20%** of revenue.

Citizen integrators hook up new systems without IT help, and system-specific AI tools only make things worse.

The need: a **multidirectional sync** that manages companies' increasingly distributed data where it's already stored.

Gartner has called it: By 2022, **30%** of new integration projects will rely on AI technology.

Security



## TABLE OF CONTENTS

**1.** The Catastrophic Cost of  
Bad Data

---

**2.** The Added Threat of  
System-Specific AI

---

**3.** How Companies Deal With  
Bad Data Today

---

**4.** The Need for a  
Smarter Sync

---

**5.** The Syncari Solution

---

**6.** Notes

---

# 1



## THE CATASTROPHIC COST OF BAD DATA

***The Economist* calls data the oil of the 21st century and considering the quality at most companies, one has to imagine it laden with gunk.**

Companies spend \$260 billion on analytics and yet only 55 percent of their data is available and only a fraction is suitable for analysis.<sup>3,4</sup> In a study of 75 enterprises by MIT, only 3 percent of data sets were within an acceptable range and 50 percent contained critical errors.<sup>5</sup>

Not only does bad data lead to wrong decisions, it also encumbers teams who spend as much as 50 percent of their time dealing with mundane data quality issues.<sup>6</sup> Thomas Redman, the author of *Data Driven*, calls this the accommodation problem: Wherever there are errors, employees spend valuable time confirming with other sources and often neglect to correct it at the origin.

# BAD DATA COSTS COMPANIES 15 TO 20 PERCENT OF REVENUE IN MYRIAD AND OFTEN COMPOUND WAYS:



**Marketing campaigns underperform.** Marketing teams build inaccurate segments, launch ineffective campaigns, run the risk of being blacklisted by ISPs, and spend an average of 800 hours per year cleaning lists.<sup>7</sup>



**Sales can't sell.** Salespeople spend too much time building lists and verifying emails and too little time actually selling. Today's salespeople spend 64 percent of their time, or 900 hours per year per rep, on non-sales activities.<sup>8</sup>



**Business teams accumulate risk.** Fifty-five percent of data is dark—information the business doesn't even know about, or can't access—which exposes them to breaches and fines.<sup>9</sup> By 2021, twenty-five percent of organizations with public APIs will have discontinued them due to security incidents, reports Gartner.<sup>10</sup>



*“Nothing wrecks sales productivity like hiring smart people and having them build lists.”*

Scott Edmonds, Sales Advisor



**Product teams build features that fail.** The number one reason products fail is a lack of insight into the target audience or market.<sup>11</sup>



**HR over-staffs and can't retain talent.** HR teams over-hire based on inaccurate revenue projections, hire for the wrong roles, and can't keep good employees from quitting.



**Customer success teams upset customers.** Companies damage customer loyalty and increase churn with redundant outreach to duplicate records, billing errors, reactive support, and shaky handoffs.

## AMIDST ALL OF THIS

Teams continue to buy third-party data to cleanse and update existing records to the tune of \$19 billion per year. What companies can't seem to maintain in-house, they buy at great expense, while the underlying issues remain.<sup>12</sup>

# \$19 billion

Spent on third-party data

# 60-73%

Of data is never used

## THE PROBLEM? CLOUD CONNECTOR CHAOS

The issue is the same legacy connections the team and I worked so hard to avoid at Marketo.

Businesses still connect SaaS applications with technologies designed during the dot-com crash without innovating on them. Most companies have 12 to 15 apps plugged into their CRM, all of which use APIs that treat everything like a point-to-point connection, and every system like a silo, with no awareness of the overall data needs of the business. APIs can be smart, but the ways companies often use them are dumb.

Even if you keep most of your software purchases within one suite or "cloud" of tools, you still have data silos. Many marketing, HR, and back-office "clouds" are just marketing fluff. They're groups of recently acquired companies that are insufficiently integrated to share data. Even if all your business units use the same CRM, there can be silos. Each organization often has its own disconnected instance, which prevents them from sharing data.

## EACH SAAS VENDOR CONNECTS ITS APP TO OTHERS IN ITS OWN UNIQUE WAY

If we stick with the data as oil analogy, your API-connected data ecosystem is an oil refinery where every pipe was built by different teams at different times. Some used the metric system, some imperial, some both. Some left instructions in Mandarin, others left none. Some pipes flow one way, but not the other. Some enrich the data. Others leak. The resulting data is sludge, and the problem is growing worse.

.....

*"Integration vendors are facing challenges in reconciling the semantic and syntactic differences in the thousands of APIs generated by SaaS and privately developed services."*

Gartner

.....



## THE COMPLEXITY OF DATA ECOSYSTEMS IS RISING EXPONENTIALLY

As the number of SaaS apps used by each company increases, the complexity of these ecosystems rises exponentially. This is known as the N-squared problem, or Metcalfe's Law: When you add nodes to a network, the network grows exponentially more complex to manage. Consider the figure below.

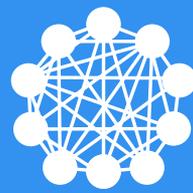
Network connections =  $n(n-1)/2$



2 applications  
1 connection



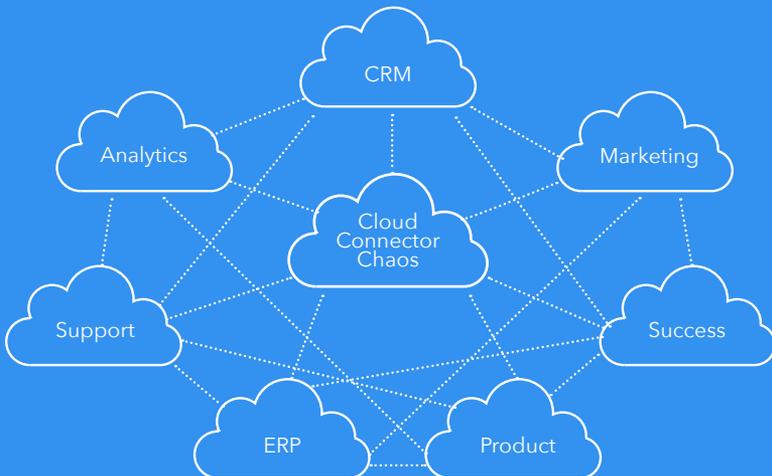
5 applications  
10 connections



10 applications  
45 connections

When you have two nodes in a network, there's one connector. When you jump to five nodes, you suddenly have ten connections. When you have 14 nodes—the average number of apps companies hook into their CRM—you have 91.

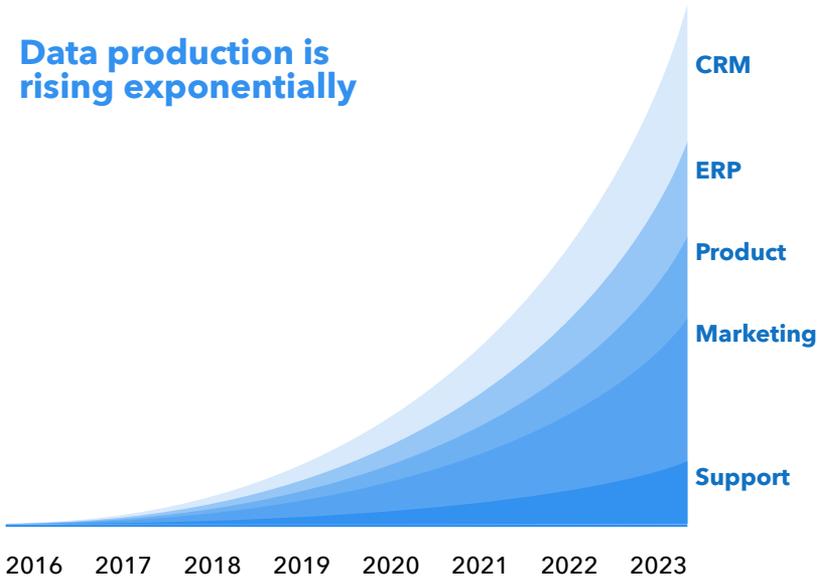
Companies don't connect every system to every other system, but the principle holds. Lots of data is lodged in places other than the CRM, or exists in duplicate across multiple systems. Good data is lost, overwritten, altered, omitted, expunged, and decayed.



## IT GETS WORSE

Lots of business units now buy and launch cloud applications without IT's involvement. In an effort to build friendlier UIs to appease these "citizen integrators," many vendors now allow end-users to rapidly reconfigure software settings and effectively change how the system stores and uses data. This means that at each of those endpoints, there are individuals like sales managers, accounts payable specialists, and engineers flipping switches and rewiring the company's data architecture.

And, rather than making data quality better, the field of AI is rapidly making things worse.



# 2



## THE ADDED THREAT OF SYSTEM-SPECIFIC AI

System-specific machine learning algorithms are making all the bad things happen faster. AI is increasingly a feature of SaaS offerings, from Salesforce Einstein to NetSuite Intelligent Cloud, and spending on AI in software will reach \$98 billion by 2023.<sup>13</sup> Forty-five percent of businesses use some form of it.

Each algorithm is specific to the system and so prioritizes data quality within its own fiefdom without concern for others, constantly optimizing based on an incomplete picture of the business.

.....

*“Automation applied to an inefficient process will magnify the inefficiency.”*

**Bill Gates**

.....

## HOW AI GOES AWRY

- **The customer support AI** “corrects” contract start dates to the European format based on the convention at headquarters.
- **The sales AI** updates lead scores based on prospect engagement, triggering a flood of emails from marketing.
- **The marketing AI** overwrites an entire list of leads’ correct phone numbers with the generic 800 numbers drawn from a new data enrichment technology.
- **The analytics AI** reformats all activity data, making it unusable to the product team.
- **The marketing AI** changes the billing contacts for finance.

## SYSTEM-SPECIFIC AI POSES SUCH A DIRE THREAT BECAUSE:

- **Each point solution is a silo.** Each AI optimizes for something different without coordinating with others, causing central data quality to decay, fast.
- **System-specific AIs are features, not products.** Not every company has the machine learning expertise or quantity of data necessary to build useful, enterprise-grade AI systems, but many do anyway.
- **Errors aren’t easily reversed.** Point solution algorithms learn, improve, and take autonomous action. But if the action is an error, teams often don’t have the tools to revert.

Data in an ecosystem with multiple system-specific AIs decays at blistering speed, and unlike human operators, AI never sleeps. Does your team have a plan?



# 3

## HOW COMPANIES DEAL WITH BAD DATA TODAY

The data quality problem is so imposing that not all companies are contemplating a fix. It's the status quo in many businesses, and its effects seem so specific to each team that it's difficult to see it as the organizational problem that it is. "New technologies and approaches are available," writes Gartner, "yet many stick to their proven practices."

**Those doing something are doing some combination of:**

- A.** Buying data to fill the void
- B.** Team-specific initiatives
- C.** Customer data platform (CDP)
- D.** Master data management (MDM)

.....

*“Men and nations behave wisely, but only after they’ve exhausted all other resources.”*

Abba Eban

.....

## A. BUYING DATA TO FILL THE VOID

US firms spent \$19 billion on third-party supplementary audience data in 2018, up 18 percent from the year before.<sup>14</sup> That’s not including data cleansing, integration, and hygiene services. There were 1,120 data vendors on the MarTech Landscape chart in 2019.

## B. TEAM SPECIFIC INITIATIVES

Individual business units or team leaders either purchase new data, implement new data procedures, or occasionally cordon their data off so other business units can’t foul it up. If they pay for third-party data and data cleansing or appending services, they become dependent on frequent refreshes.

## C. CHASING THE SPECTER OF CENTRALIZED DATA MANAGEMENT

Customer data platforms (CDPs) have become numerous enough to warrant their own Forrester New Wave report, but are built on a flawed premise of centralized storage.<sup>15</sup> CDPs move all the data to one central location, from which it isn’t easily moved back and accessed by the end users to whom it’s valuable. If I’m a marketer, it doesn’t matter to me that our email addresses are correct in the CDP if I send emails from the marketing platform. And if I want to build a bidirectional sync, I quickly run up against API call limits and excessive latency.

Moving data back and forth from a central location only exacerbates the N-squared problem covered in Chapter One whereby more connections lead to exponentially more complex problems.

## D. STRUGGLING WITH MASTER DATA MANAGEMENT

Similar to CDP, master data management software (MDM) tries to centralize data where it can be cleaned, transformed, and redistributed. It's been a topic of conversation for over a decade, but the concept suffers serious limitations:

- **Storage is expensive and slow.** Many MDM solutions try to capture all data in their original format to sort later, but storage is both more difficult and more expensive than the pioneers of the 2010s imagined. "There was a rush to capture all data and sort it all later," says Rob Zare, Senior Director of Product Management at Salesforce. "But most of that data is garbage. People collected petabytes of proverbial cat pictures." As a result, companies are growing disillusioned with Hadoop and non-relational data storage, and the sector is folding. Hortonworks and Cloudera have merged and at the time of writing, MapR is near bankruptcy.
- **Central control is problematic.** MDM initiatives fall under IT, which often doesn't have the luxury of spending enough time with the various lines of business to know what good data means to end-users. This also makes it slow to adapt as data and data sources change.
- **Requires strong executive sponsorship and change management.** Which is to say, there are very few instances of it working in practice.

All this has led me to believe that to truly tackle the data quality issue, companies need and want a system that borrows from each approach, but addresses the issue at its source.



*"Most businesspeople readily admit that data quality is a problem, but claim it is the province of IT"*

Thomas Redman



# 4

## THE NEED FOR A SMARTER SYNC

Every company that relies on data is facing increased challenges. They're connecting more applications to their CRM which erodes data quality at the center, while end-users and system-specific AI erode data at its edges.

To avoid things getting worse, companies need a system that:

- Understands which system has authority for which data
- Manages data where it exists without trying to store it centrally
- Optimizes for data quality, not connections
- Learns and improves data quality over time

Such a system would need to play air traffic controller to maintain data

quality, not just data connection. It would need a schema for understanding where data goes, but also why, and for whose benefit. It would need to understand and safeguard the purpose of the data to the business. I've outlined this in what I call the data quality maturity model.

.....

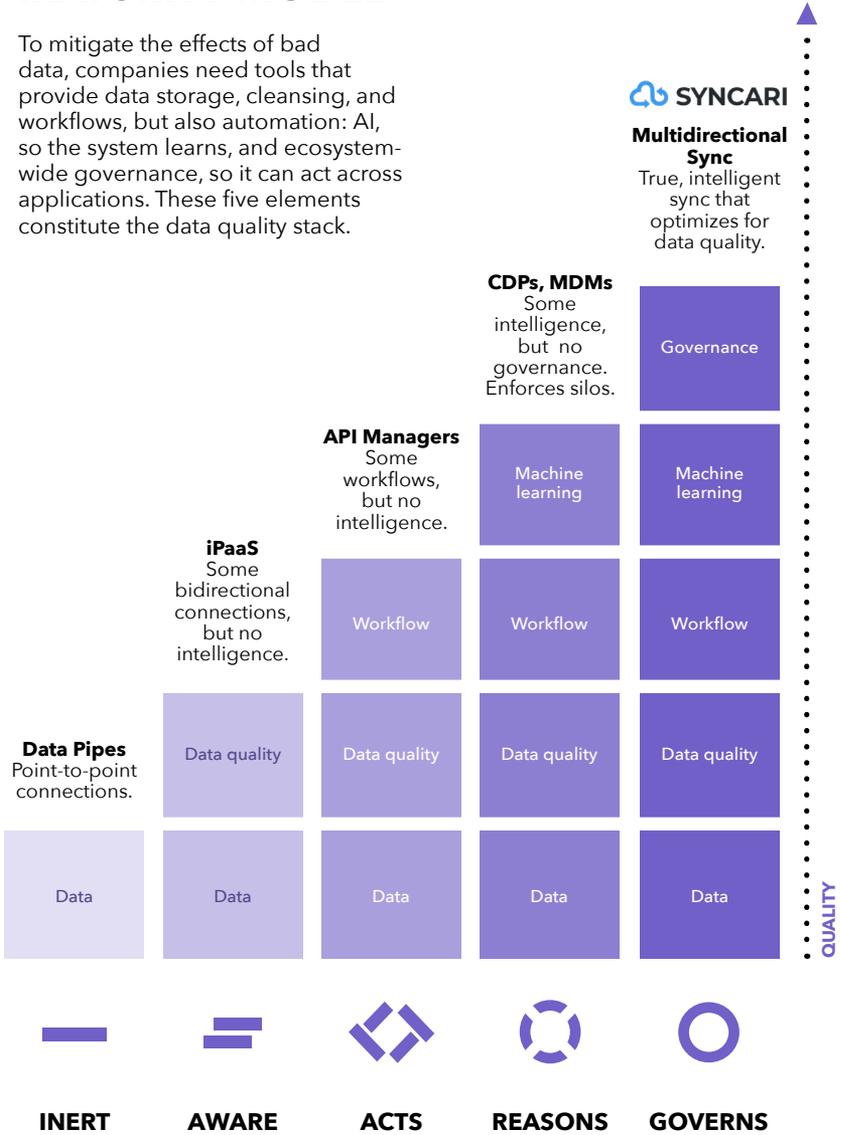
*"What I want is to make sure that good data isn't corrupted by bad data from a system that has no right updating that field"*

**Rob Zare, Salesforce**

.....

# THE DATA QUALITY MATURITY MODEL

To mitigate the effects of bad data, companies need tools that provide data storage, cleansing, and workflows, but also automation: AI, so the system learns, and ecosystem-wide governance, so it can act across applications. These five elements constitute the data quality stack.

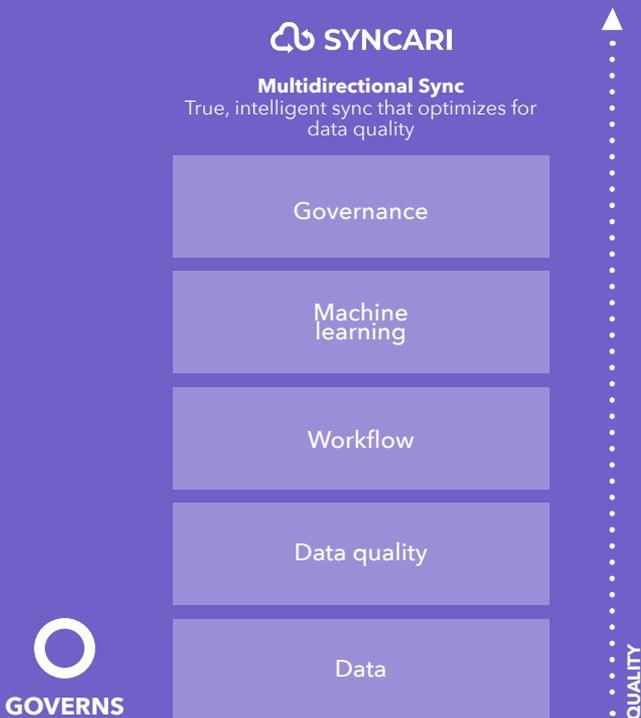


# ASCENDING THE MODEL LEADS TO DATA QUALITY

The higher up your business ascends on the model, the more complete your data stack, and the more easily you can mitigate the effects of bad data. At the Governs stage, the system managing your data ecosystem actually improves data quality over time. I call this a true multidirectional sync. Companies at this stage enjoy greater gains for every business unit:

- Marketing campaigns outperform and build more pipeline
- Sales teams spend more time selling and close more deals
- Product teams build more successful products
- Business teams reduce risk
- Customers renew and expand
- Companies spend less on third-party data

We're talking about clawing back part of the 15 to 20 percent of your company's revenue that's lost each year.



## TODAY, NOBODY OFFERS THE COMPLETE “GOVERNS” PACKAGE

No vendor offers a complete solution, or will be around long enough to. By 2023, up to two-thirds of existing iPaaS vendors will merge, be acquired, or exit the market, reports Gartner, and the iPaaS market is bifurcating between general-purpose applications and domain or vertical-specific ones.<sup>16</sup>

Today, many of these vendors are launching AI offerings. There’s Dell Boomi Suggest, Informatica’s CLAIRE, SAP’s integration content advisor, SnapLogic’s Iris AI, and Workato’s Workbot. But none of these features have AI at their core, and still treat connections as point-to-point, and data as just data. Some are simply SDKs for building AI that require extensive client-side developer resources to deploy. None offer code-free interfaces or tools that citizen integrators can actually take advantage of.

 **Data pipes simply** move data back and forth without intelligence.  
**Examples:** Dell Boomi, MuleSoft, IBM App Connect

 **Integration platforms (iPaaS)** move data back and forth with some conditional logic. Also known as cloud connectors.  
**Examples:** Zapier, Oracle Data Manager

 **API managers** simply ensure API conventions are met and documented.  
**Examples:** Mulesoft, Microsoft Azure API Management

 **CDPs** offer business logic, but introduce complexity, cost, and errors.  
**Examples:** Blueshift, Segment

 **MDMs** but are impractically expensive, difficult to implement, and lack business logic.  
**Examples:** SAP Master Data Governance, IBM MDM solutions, Orchestra Networks EBX

## SHORT-TERM RECOMMENDATIONS

What can a business do? In the near term, you can restructure your team to better manage the influx of data and point-to-point connections.

**Create the role of API manager.** An API manager's job is to assemble an integration strategy team (IST) to invest in API planning, design, and testing, and to help citizen integrators accomplish their goals while doing less harm.

**Invest in API lifecycle management.** Your team needs to document its existing data flows, mappings, and configurations, and investigate each business unit's integration needs.

**Adopt AI integration systems.** AI integration systems may help combat the damage done by system-specific AIs by automating some processes, like elastic healing and self-scaling. Some systems can make "next step" recommendations when developers are planning integration flows, and provide intelligence into how data is used throughout the company.

**Don't overcomplicate things.** The last thing your data ecosystem needs are more systems. Investigate ways to have less.

In the long term, you need technology that does this for you. Such a system would have to apply data governance globally without centralizing data, feature AI at its core, and offer a code-free interface usable by citizen integrators.

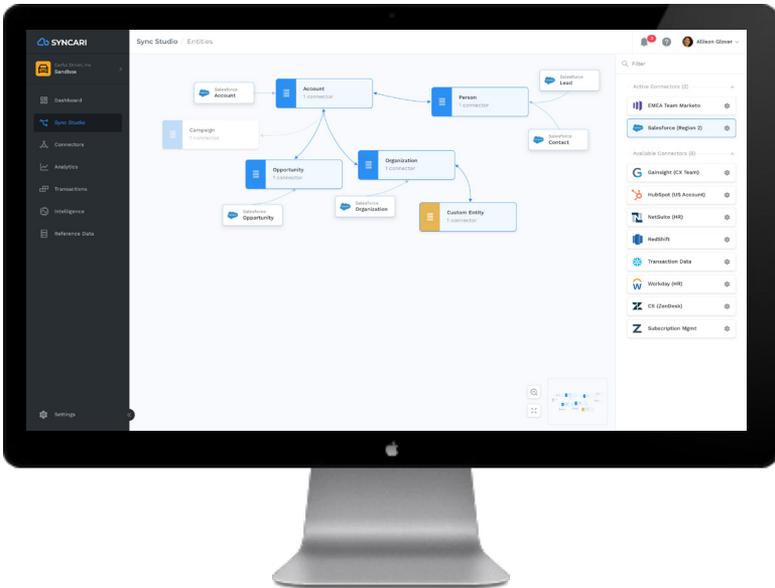
And that's exactly what we've reassembled the Marketo sync team to build.



## 5

## THE SYNCARI SOLUTION

**Let me introduce Syncari.** It serves as a distributed source of truth for all your business data. Rather than transporting it to one central location for cleansing, it monitors and manages your data where it already lies, in your CRM and across your top 12 to 15 apps.



.....

*"By 2022, 30% of new integration projects will use AI technology for faster delivery."*

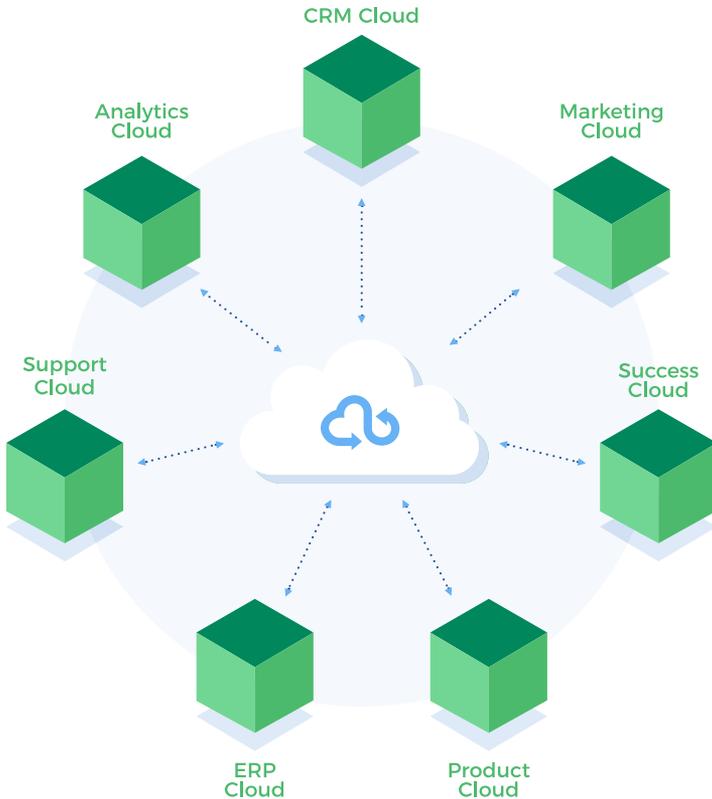
**Gartner**

.....

# SYNCARI IS A SMARTER SYNC

**It's a distributed source of truth that:**

- Creates a global data schema across all SaaS applications.
- Reduces connections (API calls) across the network.
- Provides accurate data to business-critical operational systems.
- Uses machine learning to improve data quality.
- Features a user-friendly interface that's accessible to anyone.



## THE THEORY BEHIND IT IS SIMPLE:

Your data ecosystem needs a manager that corrects and solves for data quality, not data connections. In many cases, Syncari eliminates the need for point-to-point connections and the associated API load. It continuously:

- Ensures data quality across systems
- Empowers people to use more data
- Automatically improves data quality
- Synchronizes and unifies data structures (schemas) within connected systems
- Automates object and field mapping

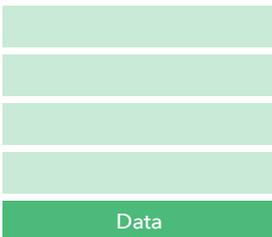
### OLD WAY

Integrations  
Connector-centric  
Simple workflow  
Ignores data quality

### NEW WAY

Synchronization  
Data-centric workflow  
AI/ML data operations  
Prioritizes data quality

### NO SOURCE OF TRUTH



### TUNABLE SOURCE OF TRUTH



# IT'S A SINGLE SOURCE OF TUNABLE TRUTH

Syncari is the long-sought-after single source of data truth. It's distributed, it's tunable, and it's a way to ensure you can deliver on your customers' expectations of a high-touch experience.

That means there's no CDP, data warehouse, or additional aggregation required. There's no need for data scientists or to architect complex cross-system data calculations. And there's no need to lock your ecosystem down to stop citizen integrators. There's just data quality, everywhere, manageable from one place, effective wherever your data already exists.

Syncari is the answer to the jumble of API integrations and the \$3.1 trillion snarl of cloud connectors, and just what you need to future-proof your business.

**Easily integrates SaaS applications:** With zero working knowledge, someone can connect two or more systems in minutes. Syncari knows exactly what to map, including custom objects.

**Multi-SaaS cloud synchronization:** Syncari keeps individual data points clean and consistent throughout your connected systems, with bi-directional syncing that reduces API calls.

**Distributed data quality:** Anyone can build models to spell-check, augment, and govern data without complicated configurations or code. (This is especially useful for predictive scoring, sentiment, and customer engagement.)

**Make your valuable data visible:** Syncari helps unlock your dark data and make sure it's correct and available to users and systems.

**Enforces data governance:** Syncari automatically deduplicates and resolves data disputes throughout your system, maintaining quality and consistency as a byproduct to tracking it.



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

## CONCLUSION

It's been more than a decade since the Marketo team squeezed into that conference room, but part of that team is gathering again, just up the road, to talk about what users don't want. Leaders, salespeople, marketers, analysts, and managers don't want to be spending 50 percent of their time dealing with mundane quality issues. They don't want to lose 15 to 20 percent of revenue each year. They just want their data to be accurate so they can do their job, so we built Syncari for maximum convenience. Last time it changed an entire market. This time, it's going to change the world.



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

## ABOUT THE AUTHOR



**Nick Bonfiglio**

CEO, Syncari

Nick has over 28 years of experience in product development, information technology, and customer experience. Prior to Syncari, Nick was CEO and Founder of Aprinsic (now Gainsight PX) where he wrote *Mastering Product Experience in SaaS*, a book about product-led go-to-market strategies. Prior to that, he was EVP of Global Product for Marketo and responsible for engineering, product management, cloud operations, and technical support.

**Follow him on Twitter at [@Nick\\_Bonfiglio](https://twitter.com/Nick_Bonfiglio).**

# SOURCES

1. IBM <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
2. MIT Sloan Review <https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/>
3. IDC <https://www.idc.com/getdoc.jsp?containerId=prUS44215218>
4. Splunk <https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>
5. Harvard Business Review <https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards>
6. Redman, Thomas. Data Driven <https://hbr.org/product/data-driven-profiting-from-your-most-important-bus/an/1912-HBK-ENG>
7. ReachForce <https://blog.reachforce.com/blog/the-time-spent-on-routine-tasks-how-much-time-is-your-team-wasting>
8. Forbes <https://www.forbes.com/sites/kenkroogue/2018/01/10/why-sales-reps-spend-less-than-36-of-time-selling-and-less-than-18-in-crm/#4ee28edbb998>
9. Splunk <https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>
10. Gartner <https://www.gartner.com/en/documents/3898763/predicts-2019-democratization-of-it-requires-different-s>
11. CB Insights <https://www.cbinsights.com/research/startup-failure-reasons-top/>
12. IAB <https://www.iab.com/news/2018-state-of-data-report/>
13. IDC <https://www.businesswire.com/news/home/20190904005570/en/Worldwide-Spending-Artificial-Intelligence-Systems-98-Billion>
14. IAB <https://www.iab.com/news/2018-state-of-data-report/>
15. Forrester <https://www.forrester.com/report/The+Forrester+New+Wave+B2B+Customer+Data+Platforms+Q2+2019/-/E-RES143852>
16. Gartner <https://www.gartner.com/en/documents/3898763/predicts-2019-democratization-of-it-requires-different-s>

**Learn about the team that helped produce the book at [findaway.media](http://findaway.media)**

