The Marketing Data Lake

How a new generation of marketing analytics (and a Big Data mindset) changes... **absolutely everything.**

Franz Aman & Anish Jariwala

The Marketing Data Lake

How a new generation of marketing analytics (and a Big Data mindset) changes... **absolutely everything.**

Franz Aman & Anish Jariwala

“No company in its right mind tries to sell to everyone.”
Philip Kotler

Contents

Let’s fix marketing

The two of us – Franz and Anish – have been in the marketing game for over thirty years. And for all but the last few of those years, we’ve held a growing suspicion – a conviction really – that marketing is broken. Badly broken. For decades, marketing departments have not been able to do even the most basic things that every other department has always done. We couldn’t prove beyond reasonable doubt that the things we did actually worked. That they were worth the investment. We couldn’t make truly informed, confident decisions about where, how and when to spend our budgets. We couldn’t defend our decisions from even the feeblest attacks. Listening to Marketers at CMO conferences we knew that we were not alone. Over and over,

we heard complaints about how Marketing was not valued or appreciated. About the short-lived tenure of the average CMO. Or the endless fights about who created pipeline and whether it was really converting… ___ Whenever a new CEO came into a company, the CMO job always seemed to be the first to go (and the first act of every new CMO was to fire the agency and rebrand the company). ___ Bottom line, we marketers have never really justified our existence. ___ And it hurt. ___ But over the last three years, our small team at Informatica has discovered a better way. ___ Disgusted by the wasteful, opaque, unaccountable marketing machine most companies are saddled with, we set out to build a better one. ___ We built it using readily available tools and technologies. ___ For less than the cost of a crappy banner campaign. ___ And we did it in sixty days. ___ Our new machine turned our marketing upside down, transforming it from the poor relation of the 'grown-up' departments into a powerful, accountable, data-driven contributor to our company's success. ___ For the first time, we could connect the dots between all of our marketing activities, end-to-end, to see which ones actually contributed to revenue. This view ACROSS platforms, silos and channels changed everything. ___ This is what marketing was always meant to be, but never was. It's built using a big data approach, around a new construct called a Data Lake. ___ And you can build one too. ___ Our new marketing machine may not look all that different from the outside, but make

no mistake: we've replaced a World War II biplane with a fighter jet. A Model T with a Tesla. And the impact on our department, our company and our careers has been profound. This book is here to help you fix your own marketing and build a completely new kind of revenue-generation machine for your company, using a Data Lake to do it. If you follow our lead and apply the ideas in this book, your marketing will change forever. We promise. It's the single most important opportunity facing every marketer. It's also the most exciting and fulfilling (trust us, it's a lot more fun discussing how much revenue an incremental marketing investment will deliver than defending the existing budget). We hope you'll join us on this expedition.

Who this book is for

This book is for CMOs, marketing operations professionals and senior marketers who want to make their marketing more data-driven (by which we mean customer-driven).

It's also for the marketing operations people on the front lines of the battle against blind marketing. The people trying to connect the dots across their own marketing programs.

The book will outline the concept of the Marketing Data Lake, its benefits and the main parts of the new revenue-generation machine. CMOs who never logged into Marketo or Salesforce will get a lot from these chapters.

It's also a how-to book for operations people who spend their working days inside marketing applications. These chapters may feel a bit geeky and specific for the CMO types but they'll still have value – if only to show you the challenges that your operations people will be facing.

We're assuming you're data inclined (or at least data curious) but that you may not be particularly techie. No code snippets or Hadoop architectures here (though if that's your thing, do get in touch). We're career B2B marketers and Informatica is a B2B company, so a lot of what we talk about will have a distinctly B2B feel.

But the ideas and approaches apply just as well to consumer marketing, with a bit of adapting. So whether your company sells ball bearings, BI software or bagged bagels, do read on.

Fair warning: the approach we outline here is best for bigger companies – companies that can justify the investment required and that have enough data to work with.

If you're in a small or mid-sized company (say, under $100 million in revenues) you'll get a lot from the book but you may not be able to do everything we've done.

But if you're leading a big enterprise marketing department, this is your blueprint.

- **A wake up call for B2B**
  Ecommerce brands tend to be data-driven. The ability to see real money flowing through their websites has increased the pressure for accountability and dialed up the precision and URGENCY of their operations.

  In the B2B world, marketers are still hiding behind the difficulty of our marketing and sales processes. Long sales cycles, complex deals and multi-disciplinary buying teams have provided an excuse to be more lax about our performance metrics.

  That's no longer sustainable. Today, B2B marketing can and must be every bit as precise, transparent and accountable as the tightest ecommerce shop.

  Saying something is hard is not the same as saying it's impossible.

Some housekeeping

Brand names
Throughout this book, we refer to the marketing systems and applications that we actually use. This is not to promote these tools. We simply want to be as specific and real-world as possible.

In truth, almost everything we use could be swapped out for a competitor with little impact on the program.

If you don't use the tools we use, don't worry: it's not about the specific technology brands, it's about the data, the mindset and the processes.

Privacy
We'll use a lot of screen shots in the book, showing our own system as it really looks. But the data in each screen is not real customer data – it has been masked (shameless plug: using our own Informatica data masking tool) to protect the names, emails and phone numbers of real people.

So don't bother emailing Mr. Bill Jones. He's fictional.

“Relying solely on gut-based, experience-driven decision-making in marketing is foolish in the digital age.”
Scott Brinker, chiefmartec.com

A data-driven marketing manifesto

Whether it's already hit your cubicle or not, marketing has changed forever.

The companies paying our salaries still need precious creative snowflakes and strategists with 'great instincts'. But even more, they need marketers who value and understand data.

In the financial department, data is about your company's money. In the marketing department, data is about your CUSTOMERS. To be lax about your data is to ignore your customers.

Data is where your customers and prospects tell you everything about themselves, their relationships to your company and their engagement with your products.

To squander that resource is a cardinal sin. It is nothing short of marketing malpractice.

But most marketing departments still think that their applications – CRM, marketing automation, analytics, whatever – are capturing

and managing their data perfectly adequately. “We’re doing data,” they say, “Just look at this dashboard!”

In truth, you’re not doing data-driven marketing until you OWN your data. Until you take full responsibility for all of your data, across all of your marketing applications and channels. Until you bring that data together to clean, integrate, manage, master and secure it.

As a marketer, your data is your single greatest strategic asset. Your job – every marketer’s job – is to treat it that way.

The answer to almost every question about your strategy, tactics, offers, budget allocation, marketing mix, content and creative can be found inside your data right now.

You just have to listen.

→ **The data dividend**
A 2015 report by the Association of National Advertisers, McKinsey & Co., and GfK reflected the rising importance of data for client-side marketers:

- The ability to make data-guided decisions was cited as the most important marketing capability (83 percent of respondents).
- Using analytics to improve marketing effectiveness was the second most important strategy (80 percent), followed by analytics for understanding customer behavior (78 percent).
- But only 10 percent believed they were effective at feeding customer insights back into the organization.

Keep your eyes on the prize

This is what the Marketing Data Lake approach that we'll describe in this book will help you do:

- See which of your prospects, customers and segments are engaging with your marketing programs.
- Track their journeys across channels.
- See which channels and programs are delivering your revenue and which are not.
- Actually see your revenue travel through your pipeline, with early indicators showing any shortfalls long before they hurt (so you can take action).
- Build a constructive, collaborative and mutually accountable relationship with Sales.
- Get an account-based view of your world, instead of just a lead-centric one.
- Create an agile marketing operations team that responds quickly to the needs of the business.

Heard this all before?

Yes, we know: every marketing technology on the market (and there are thousands) make very similar claims. We assure you: few tools

can actually do these things even within a single marketing channel – and no one platform can do them in a way that covers all of your marketing activities. If an existing platform or 'marketing cloud' could do all that the modern, accountable marketer really needs, we'd be the first to sign up.

In truth, we marketers need to step up and own these problems. To own our data. And to integrate and manage everything to create a complete picture. You can't sit back and wait for someone to invent this. You need to make it happen.

The good news: it's not as hard as it sounds. It's well within your reach. You may need some new skills or tools but if you have the will, the way is clear.

Trends in marketing: a pattern emerges

The Marketing Data Lake hasn't come out of nowhere. In fact, it's a natural – or even INEVITABLE – development in marketing.

A look at all the main trends in marketing shows a distinct pattern emerging:

Digital marketing – Web, mobile, social media and email marketing are all dramatically more measurable and accountable than old-school, 'broadcast-style' marketing. We take all this for granted today, but without it, we just wouldn't have the data to fuel this new model.

Content marketing – Packaging up your expertise to help your prospects and customer do their jobs (or enjoy their lives, if you're in B2C) has proven to be the best way to attract people and engage with them. It's also the fuel of the new, data-centric marketing engine.

Big data and next-generation analytics – We now have the desire and ability to combine data – structured and

unstructured – from many sources into a single analytical framework. Big data is transforming many disciplines, but marketing, with its wide range of weird and wonderful data sources, is a natural beneficiary.

Personalization – The concept of intelligent segmentation and targeting, right down to the segment of one, has been around for decades. But only today have the ingredients all come together to make it ready to go mainstream as a marketing discipline.

The convergence of sales, marketing and customer service – As each customer-facing discipline realizes it's actually in the customer experience business, the walls between them are falling. A single Data Lake can serve all of these disciplines (as well as the Product teams), helping them improve and unify the customer experience.

DMPs and the AdTech boom – The online advertising world has been a hotbed of innovation, with sophisticated, data-driven ad servers, real-time bidding and Data Management Platforms. To us, the Data Lake is like a flexible, omnichannel DMP that lets you, the marketer, own and manage your own data (instead of leaving it to a third party).

The rise of iPaaS – As data integration becomes a strategic discipline across the enterprise, the Integration-Platform-as-a-Service concept (in which integrations are treated as re-usable patterns) is gaining traction fast. iPaaS was identified as a distinct category for the first time in the 2016 edition of the Chiefmartec Marketing Technology Landscape. Expect it to grow.

These and other trends are coming together to create a new environment – and new tools – for the modern marketer. As you'll see, the Data Lake concept sits right at the crossroads of all of these drivers, making it easier to harness the new potential.

“Without big data analytics, companies are blind and deaf, wandering out onto the Web like deer on a freeway.”
Geoffrey Moore

How it all started

Back in 2014, before we embarked on the Data Lake program, the Informatica marketing department was shattering all of our targets. Just as a sample:

> We said we'd increase traffic to our website and content by 35 percent – we doubled it.
>
> We over-achieved our pipeline goals by 42 percent.

We were proud of our achievements. We'd come a long way from a department that couldn't even measure these things, much less move them.

Then reality kicked in: the numbers showed that marketing-generated revenue had not climbed at the same clip as marketing-sourced pipeline.

Don't get us wrong, we had a terrific year and delivered on our projections to investors. But those 'Marketing Qualified Leads'? Turns out, they weren't as qualified as we expected.

And that pipeline? Not enough of it was converting into real revenue.

The writing was on the wall: if the numbers were correct, marketing was simply not doing enough to drive real revenue.

That left us with two options:

> We could blame Sales for failing to convert all that wonderful opportunity we'd created for them.
>
> Or we could admit that we weren't generating real, closeable opportunities despite the apparent pipeline numbers.

For us, the choice was painful but clear. We knew that our products ran circles around the competition and that our sales team is one of the best in the business. Give them the right opportunities and they'll close them.

The problem was somewhere in our marketing machine.

A commitment to change

So, as a team, we committed to solving this problem. And to find the answer, we needed to measure the effectiveness of two key marketing areas: our data and our programs. We also had to better align with sales priorities and coverage.

We had a wealth of amazing data from our marketing automation system, our CRM, and our finance system (for actual revenue data), but we needed to get to the bottom of the oldest problem in marketing: **What actually works?**

We already employed a wide range of marketing programs and tactics, including SEO, paid search and social, content marketing, events, display advertising, social media marketing, and analyst relations. We just needed to understand which ones actually helped us achieve our goals and which we needed to fix.

With marketing technology being so rich, answering these fundamental questions could not possibly be beyond our reach.

We had to crack this. And if we did, we'd be able to do something very few B2B marketing departments have ever accomplished.

That's what this journey is all about.

“Crawl, walk, run.”
Franz Aman

Let us introduce ourselves

To put the program we're about to share in context, it helps to know a little bit about us and the company we work for.

Rest assured, this is not a sales pitch and nowhere in the book will we try to sell you Informatica products (though we will describe one or two that we used in our Data Lake initiative).

The company we work for

Informatica makes data management tools that help companies access, clean, integrate, manage, master and exploit their data.

We're a global $1billion+ company with lots of different products: licensed, on-premise software products and subscription-based, cloud-delivered (SaaS) products. It helps to know that because we may refer to our 'cloud products' or 'licensed products' a few times – to show how you can (and at times have to) treat different product lines differently in your analytics.

The team

Any change management consultant worth his or her fee will tell you that change depends on the orchestration of people, process, and technology.

But most transformation stories skim over the people part. That's a mistake. Because your people will unquestionably determine your success or failure.

The people living this Data Lake story – the Informatica marketing team – aren't rocket scientists. We're just career B2B marketers with real curiosity about the machinery of marketing operations and the data that fuels that machinery.

We wish we could introduce you to everyone on the team – there are more than 30 of us and not a weak link in the chain – but the core team in this particular journey is made up of three key players.

I'm Franz Aman and I'm a dataholic

My name is Franz. I'm the Senior VP of Marketing at Informatica.

I came up through tech marketing, including senior roles at Sun Microsystems, BEA, BusinessObjects, SAP, and SGI.

I guess you'd call me an early adopter type. I believe in the power of technology to automate, streamline, and accelerate any discipline – especially marketing.

Over the last 10 years, I've become increasingly data-oriented in my marketing strategies and tactics.

As you'll see, I'm also a curious combination of impatience and pragmatic realism (might be a German thing). I'm impatient about letting the status quo stop us from doing what we need to do – and I like to strive for improvements all the time. But I'm pragmatic

enough to take it one step at a time. This is the real world after all.

So that's me.

I'm leading our big data adventure, but it could never have happened without Anish and Laura, two smart, focused, dedicated professionals I'm honored to work with.

Anish Jariwala

Marketing Performance Manager

Anish leads our marketing analysis and campaign performance effort.

He spent his whole career doing data-driven marketing (at places like Dell and Autodesk), so he doesn't really do 'hunches' and 'instinct'. He does evidence. Most recently, he worked at Platfora, a big data discovery startup, and Marketo (both of which came in handy).

Anish knows Marketo and Adobe Analytics inside-out – a hugely valuable skillset for creating an agile marketing operations function. He and I are joined at the hip for this journey.

Laura Wang

Head of Marketing Programs and Operations

Laura is our VP of Marketing Programs and Operations. She thinks things through, then gets things done – a killer combination.

Laura worked with Franz back at SAP, SGI, Sun, and BEA and also worked at Business Objects, so her whole career has been in tech marketing, much of it in operations and program management. (They've worked together for longer than she's known her husband, Chris!)

We see Laura as a pioneer. A new kind of marketer in a new world. As marketing operations grows as a discipline, Laura is inventing the job as she goes along.

She does it by knowing how to operationalize marketing processes – understanding the desired business outcome, the technology, and the people – and then making the whole thing work at scale.

Every new B2B marketing department needs a Laura and an Anish (if you can find one).

Our friends in Sales and IT

As a team, we spend A LOT of time with our colleagues in Sales and IT – and there's no way we could have made the progress we've seen without their active participation.

We'll hear from two of these colleagues later on but we wanted to give them a shout-out here.

We're all in this together.

The people who inspire us

There are a few people who aren't strictly on the team but are very worthy of mention. They're the people who inspired us by generously sharing their ideas about what data-driven marketing looks like:

> **Scott Brinker** → The editor of ChiefMarTec.com and program chair of the MarTech Conference. He's also co-founder and CTO of ion interactive. Scott is way out in front of this marketing technology thing and he's always willing to share what he knows.

Megan Heuer → VP and Group Director at SiriusDecisions. Megan knows more about sales and marketing than the average 10 experts put together. And she's always super-articulate about the best ways to align marketing and sales.

Craig Rosenberg → Editor of the Funnelholic and co-founder and Chief Analyst at TOPO research. Among other things, Craig is a pioneer of Account Based Marketing (ABM), which is a big part of what we do at Informatica.

Jon Miller → Founder of Marketo and now of Engagio, an Account Based Marketing and selling platform. Jon has always evangelized the kind of data-driven marketing we adhere to. And now he's turning his sights on ABM.

Tom Davenport → Distinguished Professor of Information Technology and Management at Babson College. Tom pioneered the concept of "competing on analytics" in 2006 and has extended his research into analytics and big data with questions like, what will happen to humans when smart machines make important decisions?

Laura Ramos → Forrester Vice President and Principal Analyst, is a leading expert in B2B marketing. Laura's research addresses the skills, technology, process, and customer experience concerns that top B2B marketers need to understand.

Dan Vesset → Program Vice President of IDC's Business Analytics research is a long-time expert on business intelligence. Much of his research is now focused on big data.

Douglas Karr → Founder of the Marketing Tech Blog and an intelligent analyst of all things martech.

Ankush Gupta → A tech marketing veteran and Editor in Chief of Martech Advisor, a rich resource for data-driven marketers.

Adam Greco → A pioneer of web analytics and Senior Partner at Analytics Demystified, the analytics consulting firm. He's also the author of the first-ever book on Adobe Analytics, Adobe Site Catalyts: An Insider's Guide.

“If you aren’t getting the marketing metrics you need, it’s probably because you haven’t made them a priority.”
Jon Miller, Engagio

What we did – the very busy executive summary

This book is about our own transformation from old-school marketing to a new, data-driven model based around a holistic view of our prospect and customer journeys. That transformation itself was quite simple (some of the modeling, data plumbing and analytics were not) and here's what we did:

- We took all of our data from our web analytics, marketing automation and CRM systems, cleaned it up and put it all into a single Data Lake.

- We enriched that data using outside sources and ran predictive analytics against it to spot our best revenue opportunities.

- We then put data analysis and visualization tools on top of it so that we could run new reports, queries and drill downs for whatever use case we chose.

- We enlisted our key stakeholders along the way so we served them better and continually improved our revenue machine.

That's it. The rest of the book will explain what this really meant in practice, tracking our own journey and sharing the inside view of our transformation.

“Scientists prefer messy data to losing data.”
Martin Fowler, Chief Scientist, ThoughtWorks

Introducing the Data Lake

As marketers, it's good to be suspicious of anyone who claims that technology is the answer to your problems. It never is.

Making your marketing operations more agile and accountable is no different. Technology alone can't do it – it takes a well-orchestrated combination of people, process and technology.

Having said that, sometimes a new generation of technology comes along and releases enormous new potential, enabling new processes and ideas that were held back by old, legacy technologies.

The Data Lake is just this kind of major, discontinuous innovation and it makes possible order-of-magnitude improvements in analytical agility, transparency and business responsiveness – especially for the kind of data marketers deal with.

So if you're not technical at all, bear with us. This part is important.

So what is a Data Lake?

A Data Lake is a data repository that lets you store and process all your data, from multiple sources, in its native format without having to pre-structure it.

Sounds simple – but the implications for marketing are enormous.

Does the Data Lake replace Data Warehouses?

Actually, the Data Lake complements traditional data warehouses, bringing in new possibilities that are especially relevant for marketers.

Here's a quick comparison:

	Data Warehouse	**Data Lake**
Data types	Structured.	Structured, unstructured or multi-structured.
Database schema	Schema-on-write.	Schema-on-read.
Cost	Expensive storage or appliances.	Low-cost storage, commodity components.
Ideal for	Penny-perfect, super-secure analysis like financial reporting.	Agile analytics and decision-making for marketing, product development, support, etc..
Agility	Bringing in additional data is often complex and requires an IT project. Answering new questions different from what the warehouse was built for may require restructuring.	As long as the data is already loaded it's easy to add new reports and queries, combining the data in any way.

A quick look at the main ideas from the table:

Structured vs unstructured data
Structured data comes from places like transaction systems or financial systems, with neat, clearly delineated and labeled fields. Unstructured data might come from logs (like clickstream data) or streaming feeds (like Twitter). Multi-structured data means data from different sources, each with its own structure.

Data schema
A data schema is simply the blueprint that shows how your data is organized. Schema-on-write means the data architect or data warehouse administrator has to define a structure – with rows, columns, dimensions, cubes – first, before pouring in the data. Schema-on-read (also known as late binding) means you can bring the data in without pre-defining how it's organized – and organize it when you need to use it (read it).

This distinction is profound and is the key reason that the Data Lake is so powerful for marketing analytics and operations.

This is so important, we'll give it a section of its own, below.

Cost
Data warehouses are expensive to design, build and maintain. That means it can be prohibitive to store all the data you may one day need. Data Lakes allow for cheaper storage, so you can keep everything – even data you're not sure you'll ever use.

Ideal for
We wouldn't recommend using only a Data Lake for reporting financials to the board or to Wall Street. For that, you need penny-perfect precision. But we wouldn't use a warehouse for the agile, on-the-fly analytics that marketers live and die by.

Agility
If you want to add new data or run a new report that isn't supported by the way the data is structured in your data warehouse, you'll have to join the line outside the IT department

↓
Pour it in the Lake
CRM data
Marketing automation data
Web analytics data
Ecommerce data
Transaction and POS data
Social media data
Third-party data
Any kind of data!

↓
Schema-on-read changes everything, freeing you, the marketer, to collect all your data, save it forever and use it however and whenever you need.

and beg for some resources. If you're lucky, they'll write some new code for you in a few weeks or months. With a Data Lake and the right data management tools, a non-technical user can ask a new query or generate a new report without bothering IT. As we'll see, that's a very big deal.

"The need for increased agility and accessibility for data analysis is the primary driver for Data Lakes."
Andrew White, Gartner

The bottom line
Most large enterprises will deploy both data warehouses and lakes – sometimes linked – for different purposes. In my experience, the structured data warehouse is rarely agile enough to really help marketers. But that doesn't mean it isn't a hugely powerful tool for many other enterprise uses.

There's also nothing wrong with leveraging the great work that's already been done in the warehouse and bridging the two worlds. For example, rather than rebuilding historical pipeline to create year-on-year comparisons, you can mesh that historical data from the warehouse with data for the current quarter in the Data Lake. Modern analytics tools let you access both.

Where Hadoop (and Spark) come into all this

If you eavesdrop on discussions about big data, analytics or Data Lakes, you'll often hear the words Hadoop and Spark.

Hadoop is an open source framework for distributed storage and processing of very large data sets. It's fast, low-cost and ideally

suited as the home for your Data Lake. And there are lots of open source and vendor tools available to help you manage both the Hadoop environment and the data in it.

Spark is 'a fast and general engine for large-scale data processing' that's compatible with Hadoop. It performs a lot better for analytical workloads than classic MapReduce on Hadoop. For the Data Scientists among us (people whose first love was the Bat Cave of statistical computing known as 'R') there is SparkR, which lets you connect your R program to a Spark cluster.

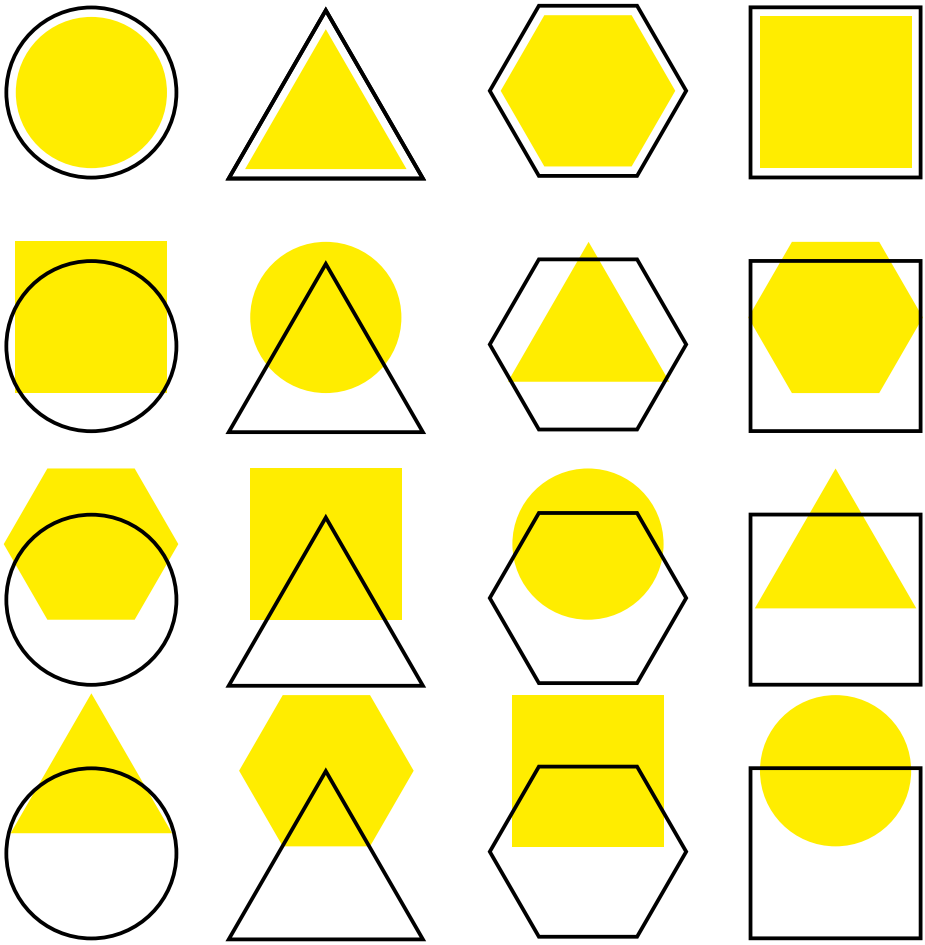
We don't have to go into these two new technologies here. The important thing is that they've made Marketing Data Lakes possible, affordable and agile.

Where do you put all this data?

Some data analytics teams keep their Hadoop clusters and Data Lakes in the cloud, in ecosystems like Amazon Web Services (AWS) or Microsoft Azure. Others live on servers (physical or virtual) in an owned or hosted data center.

We initially used AWS for our proof of concept but soon moved it to our hosted servers under our IT department's stewardship – for a variety of reasons.

The public cloud option gives you a fast, low-cost way to start, then offers limitless scalability as you bring in more and more data. But the costs do scale up too. So if you've got a lot of capacity in your data center anyway, this might prove a good option. Best to work with your Cloud and Enterprise IT architects.



Schema-on-read: a very big deal

Schema-on-read sounds like a small, technical database feature. In reality, it represents a paradigm shift in analytics – and is the key driver of value in our Data Lake program.

Being able to collect all our data without deciding in advance how to structure it means we can keep coming back to our Data Lake with new questions, new reports and new kinds of analysis. That's the essence of agility.

Marketing operations changes far too quickly to be able to predict every query you'll ever need to run. If we had to ask IT to re-engineer our data warehouse or hand code a new report every time we needed to try a use case, we wouldn't be writing this book – we'd still be at the very start of our journey.

Square pegs, no holes

A simple metaphor for the schema-on-write approach of traditional data warehouses is the child's toy that matches square pegs with square holes, round pegs with round holes, etc.

In this model, the only data you can store in this board is data that fits a predefined shape – there's no room for the unexpected.

Of course, with a structured system like this, it's very easy to count up all the blue or triangle-shaped or blue AND triangle-shaped data elements.

But what happens if you want to analyze star-shaped data? Or hexagons? You'd have to call the IT board-maker and have the base re-made for your awkward new shapes.

It's far easier to put all the shapes into one big bowl and then discover the shapes you care about and build the schema-on-read for what you need.

Remember: the Data Lake doesn't say that you never structure your data. It just lets you structure your data for each analysis and any use case, AS NEEDED.

If you try to trace back the power of next-generation analytics and look for its source, we'd say it comes from this 'little technical feature' called schema-on-read.

For the first time, marketers have the agility we need to turn a chaotic, multi-structured data environment into a flexible, manageable, STRATEGIC asset.

We could not have developed our Data Lake program if we were forced to use a data warehouse. Marketing data is just too varied and unpredictable – and marketers never know what new data source might come along to add new insight to our work.

“Show me the money.”
From 'Jerry Maguire'

The business case for a ‘big data’ approach

Let’s face it, ‘big data’ has become one of the most hyped concepts in a technology market famous for over-promise.

So before going into the ‘how’ of Data Lakes and big data marketing, let’s talk about the ‘why’.

We’ll blow away some of the smoke, explain what big data marketing really is, and show that it’s far more important than just an enterprise tech fashion trend.

The big problem: data fragmentation

The elephant in the room for every B2B marketing department today, (and most B2C departments too), is data fragmentation.

We all have so many marketing applications (we saw someone estimate an average of 30 for most marketing teams) – from CMS, CRM, and analytics to SEO, marketing automation, email,

personalization tools, social media streams, mobile apps, and campaign management platforms.

But here's the thing: each one of these apps BOTH CONSUMES AND GENERATES DATA.

Without a Data Lake, each of these apps lives in its own data silo. This fragmentation is the bane of every sales and marketing operation that cares about optimizing budgets and maximizing returns.

In building our business case for our big data marketing program, we assembled a lot of the problems we would be trying to solve – the symptoms of data fragmentation that we were experiencing in every data and that were holding us back.

You may want to reference the same things when you build your own big data marketing business case. But first, a word about big data.

What big data marketing is not

A lot of people think big data marketing is all about the sheer volume of data.

In truth, you can do big data marketing even if you don't have that much data. At Informatica, we're dealing with fairly modest data volumes but we're still attacking it with a big data approach.

After our first few use cases, we worked with these kinds of data volumes:

- Every day, we got about 24,000 rows of data from our Adobe Analytics, and 20,000 rows from Marketo.
- In our Data Lake, we had 593 columns for Adobe Analytics, 424 for Marketo, and 2,206 for Salesforce – 3,223 columns in all.
- In Tableau (our data visualization platform), we created far

fewer columns of data: 30 for Analytics, 14 for Marketo, and 4 for Salesforce.

– In total, we ran our big data marketing program for four months and we still only had 640GB of data in the lake.

So, yes, it's a lot of data but it's by no means the epic volumes that most people think about when they hear the term 'big data marketing'.

Never mind the size, feel the variety

If big data marketing isn't about the sheer volume of data, what is it about?

It's about the diversity of data sources and the range of data structures that this brings, including:

– Totally unstructured data – like text from Twitter tweets

– Semi-structured – like a text field in Salesforce

– Fully structured – such as data from transaction systems

The Data Lake approach is ideally suited to this kind of challenge, allowing you to bring in new data sources – no matter how they're structured and without re-engineering the warehouse to accommodate them.

For our program, we bring in all the relevant data from Adobe Analytics, Marketo, and Salesforce into our Data Lake – even columns we don't have any use for right now – and play with it later, slicing, dicing, blending, and mixing to our hearts' content.

In other words, **we can develop use cases as we go**.

We've been in B2B marketing for over three decades and this simple, technical innovation has allowed us to do things we could

only dream about in the old days, solving once-crippling problems that most marketers still accept as 'just the way things are'.

And it's hugely exciting.

The problems solved by big data marketing

Data fragmentation causes all sorts of expensive headaches for B2B marketing teams like ours, including:

No consolidated view
Data is locked inside different marketing applications, so it's hard to see that an action in channel A caused a response in channel B.

No account view
Marketo and Salesforce are lead-centric systems. They don't give you a clear view of activities within a company or the relationships within buying teams. As Account Based Marketing (ABM) becomes more and more important, this is unsustainable.

No view of product interest
Fragmentation meant we couldn't see exactly which products a given customer was interested in and that's a big deal for a multi-product company like ours. The data is actually in our possession, but it's buried in activity logs and we had to look across Salesforce, Marketo, and Adobe Analytics to stitch it together.

Little ability to personalize
Without a total, 360º view of each customer's activity and behavior, we couldn't really personalize our engagements with them consistently across all channels. So web personalization and email segmentation and offers were inconsistent and in some cases for some channels not relevant enough. Of course you can use personalization on the website, but chances are it will look very different from what's happening in email or other channels.

Difficulty changing or updating reports
We had to start a whole new IT project for every new report or analysis whenever we wanted to include an additional data field or join data organized a certain way differently. Adding clickstream data to the warehouse was on the roadmap for three years running – it never happened.

No granular view of programs
For instance, we had a single, "Web Paid + Web Earned" nurture flow and we couldn't separate them out to see how they differed (so we could treat them differently). Now we can.

Poor attribution modeling
We couldn't really answer fundamental questions like, "What works?" or "Did that campaign pay off?" (Last touch reports in Salesforce are not the answer!) That's not just crippling, it's embarrassing.

Connecting the dots all the way to revenue
We were focused on generating pipeline and that's what we incented and measured. But we left the question unanswered about which pipeline converted to revenue, and which programs contributed actual revenue. Now we calculate multi-touch attribution for revenue won by marketing channel and program.

That's why big data marketing is such a big deal

We could go on, listing more and more symptoms of data fragmentation and marketing sclerosis. But we're sure that if you're even trying to make your marketing more data-driven (which we sure hope you are) you're experiencing this kind of thing every day.

When you add up all of the problems that this brave new world of big data marketing can solve, the expression often used is 'the business case writes itself'.

But business cases don't write themselves. Marketers need to write them.

In a later chapter, we'll talk about writing your own business case. But the principle is simple: pick a use case that attacks one or more of the fragmentation problems listed above; then calculate the revenue gained, cost saved, or both, that happens when you solve the problem.


Q
&
A

Q&A with Adam Greco

Author and Senior Partner at Analytics Demystified

There isn't much about analytics that Adam Greco doesn't know. He's one of the smartest, most progressive analytics experts we know, with a real talent for clarifying and solving issues for the clients of Analytics Demystified and simplifying the world of analytics for the market in general.

Q **As marketers try to get a more holistic view of their world, what challenges are you seeing?**

Adam The biggest thing I see is that people think of their analytics as just a web marketing tool. It's great as a tool to optimize your site, but the next step is to take data from your web analytics and shoot it bi-directionally to other systems. To integrate it with Marketo or Salesforce, for instance, as you guys are doing.

Then to close the loop and use insight from further downstream to optimize web experiences.

A lot of B2B marketers use their web analytics tool simply

to understand everything that happens up until a form is filled out. Then they stop. Of course, there's a lot of value in that. But there's so much more you can do. For example, seeing which marketing campaigns lead to form completions is interesting, but seeing which campaigns lead to downstream revenue is more impactful.

Q **We're using Adobe Analytics in our program but we know a lot of people use Google Analytics too. Is there anything you'd do differently in GA?**

Adam Not really. Google Analytics has come a long way. I'd be surprised if there was much that you couldn't do in GA that you can do in the major analytics packages. Each will have its own approach but the concepts are the same.

Q **We use predictive analytics for our scoring but we know a lot of people use scoring in Marketo or Eloqua. How important is visitor or lead scoring in data-driven marketing?**

Adam It's very important. Predictive is pretty advanced, but even if you use a basic methodology – within your marketing automation tool – the idea of segmenting your leads by level of intent or engagement is powerful. You need to guide your sales and marketing efforts. You can't call everyone.

Even if you only do simple scoring in your web analytics tool, you can pass it along to your marketing automation or your CRM tool for prioritizing and closed-loop analysis. For example, I see many clients assigning scores to anonymous visitors and when they complete a form, pass that score to their sales team to determine if they are qualified enough to engage.

And you can score your entire website or any program you run, too. You can define your own global engagement metric and see how different activities affect the overall visitor engagement with your programs.

Q What are some of the low-hanging fruit for teams that are fairly early in their analytics journey?

Adam There are a lot of quick wins out there, wherever you are. For some, it might be identifying visitors who haven't yet completed a form – with tools like Demandbase or D&B (as you guys do).

This allows you to give sales people a heads-up about activity within certain accounts that might be actively in product research mode. Or you can use these insights to personalize the site experience by account or industry segment.

The key is to be proactive about figuring out who's on your site. Don't just sit back and wait for the form fill. Earn it – by delivering relevant experiences. Keep in mind that many companies make product decisions before they ever complete a form!

Q What data sources have you seen combined with web analytics and the core marketing data sources?

Adam With the Data Lake concept, there aren't really limits. I've seen data from tools like ClickTale – that use video to track every session – being brought together with conversion data or web interaction data. Imagine being able to show your sales team the exact sessions their key account prospects are having. This allows them to get into the heads of their prospects and improve their chances of winning deals.

Web analytics data is a rich source of insight when combined with other sources. In a silo, it's powerful. But break down the silos and you can uncover many synergistic effects.

Q **Is the Data Lake for everyone?**

Adam Probably not. You need a certain level of sophistication before you go as far as you guys are going. The important thing is to make sure your core applications – analytics, marketing automation and CRM – are in good shape, with solid data quality. Don't try to build a house with broken bricks.

If the sources feeding your Data Lake aren't robust – or you're using inconsistent tags across apps – you'll struggle.

But a Data Lake is the natural progression once you're bringing well-governed sources together.

“Give me six hours to chop down a tree and I will spend the first four sharpening the axe.”
Abraham Lincoln

The five foundations for big data marketing

In the introduction, we referred to a 60-day process for implementing our Data Lake and initiating our big data marketing project.

Now for a confession: our 60-day sprint was only possible because of the work we'd done in the previous 18-24 months.

This chapter will give you an overview of that foundational work and explain why it's been so important to our success with this new way of marketing.

To do that, we'll discuss the five foundations that we put together before creating our Marketing Data Lake. But keep in mind: the things we needed to assemble and the order in which we tackled them reflected our own unique situation and our legacy technology stack. You may need to start in a different place.

The important thing is to honestly assess where you are today – including the quality of your data and the state of your main marketing applications – and fix any foundational problems you have before

embarking on the more sophisticated techniques that the Data Lake enables.

The goal: total alignment

The big win we were aiming for in transforming our marketing operations foundations was a total partnership between Marketing and Sales. Not the lip-service kind, the REAL kind.

In the past, we had the typical 'throw it over the wall' model, in which Marketing created leads and opportunities in a vacuum (our 'pipeline') and threw them over to Sales to convert into revenue.

That wasn't working. Too few of the opportunities were converting to actual revenue. So Sales stopped trusting our marketing machine. That meant they wouldn't jump on everything we threw over the wall – so we stopped trusting THEM. Lose-Lose.

We knew that the new model had to be a complete and open collaboration between Marketing and Sales. After all, we REALLY ARE in this together. And that meant a totally transparent operational model: where Sales could see EVERYTHING we did to create leads and opportunities – and we could see everything they did to turn these into wins and losses.

At the start of this journey, we were nowhere near that ideal. And the state of our marketing operations – our data, applications, integrations, and processes – was a big reason.

So here's what we did...

The five foundations

Getting our data – and our marketing automation – in order
The first thing we needed to fix was our data.

We'd spent a lot of time and money building a marketing database and it just wasn't being used. That's usually a sign that it's not doing what it was meant to do.

We'd rolled out a central service model, building a data warehouse that integrated our Salesforce and Eloqua data – but we'd run up against some license limitations and found the size of the database was becoming really expensive.

So we knew we needed to re-negotiate and re-organize the data, and re-think our central warehouse.

At that time, our Eloqua contract was up for renewal and we had to move up to a big new release. Since that would mean a pretty major migration anyway, we used it as an opportunity to re-evaluate our marketing automation platform and our data strategy.

Pretty quickly, we ended up with a shortlist of two:
Eloqua and Marketo.

In the end, Marketo won because of ease of use (our non-technical marketers needed to be able to use it) and because we felt the integration with Salesforce would be easier and more robust.

The change also gave us an opportunity to clean up our processes and practices – to start with a clean slate, using all we'd learned with Eloqua – and set up a cleaner, simpler, better-integrated marketing automation platform.

> A BRIEF ASIDE HERE: we didn't want all our customers and prospects already in our database to have to re-authenticate themselves just because we'd changed marketing automation platforms.

So we asked our IT partners to write a nifty bit of JavaScript to port over all our prospect data so that our web forms would still pre-populate for known prospects. This is just one example of things you can do that are important but aren't out-of-the-box. So having someone who can crank out some JavaScript is a big benefit.

→ **What our new marketing automation instance and database provides us with:**

A more usable and approachable database with better quality data

Better and lower cost integration of Salesforce and Marketo data (sales pitch alert: this is a complete, two-way sync, including custom fields, synced using our own Data Integration product)

A single source of truth for Sales and Marketing

Mutual accountability!

Getting our website and web analytics in order

Next up was our website and analytics. We'd previously built on SDL Tridium, which was a great content management system for localizing our content all over the world but had some serious limitations (chiefly around things like personalization, ease of use, and mobile-friendliness).

We were also using Adobe Analytics but a very vanilla instance of it, which was pretty bare bones – not much better than the free version of Google Analytics (which is fine for the basics but doesn't let you do the kind of modeling and reporting we'd be needing).

We switched to the Adobe Experience Manager for our whole web stack and ramped up the analytics to where we needed them to be. We looked at lots of web platforms and 'marketing clouds'

and felt that – at the time of choosing – Adobe was way out ahead in integrating the stack, with strong personalization, responsive/mobile capabilities, and all the stuff we'd need (meta data capabilities for a robust taxonomy, tag management, etc.). For web data geeks like us, it's a great platform (and it's all Java-based and scalable).

We also needed to see a healthy developer and integrator ecosystem that would keep the platform growing and innovating. That was important to us. We don't want to find ourselves in a dead end ever again.

Next, we ramped up our analytics to do really robust visitor tracking, with advanced analysis to really track down conversions and affinities. The idea is to be able to map prospects to actual product interests on the web, leveraging our product taxonomy in the digital asset management system behind Adobe Experience Manager (for example, visiting web pages and downloading assets for a product like Master Data Management (MDM) accrues to MDM product interest in the affinity matrix).

Without a tight, well-run website, we just couldn't have pulled off the big data marketing operations machine we knew we needed.

→ **What our new analytics does for us**

Tight tracking of all visitor journeys

Hard-core web analytics to see page flows, referrers, repeat traffic, etc.

Integration with Demandbase for reverse IP mapping, tracking of firmographics like industry, company size etc.

A scalable paid media and SEO program

The machine was starting to hum. Now we needed some traffic to pour into it.

We ramped up our content marketing program (with our agency partner Velocity in the U.K.), and used it to fuel a modern digital marketing program to get people into the top of our funnel.

The strategy was pretty simple: create compelling content on issues we knew our prospects cared about, then offer that content in an ongoing, multi-channel advertising and social media program that we optimized over time.

We tried all sorts of tactics – from paid search, content display and remarketing to paid marketing programs on social channels, such as LinkedIn.

The result: pretty good traffic (about 30 percent of total web traffic) that we could then track as it moved around our world.

→ **What our media program did for us**

Generated a high volume of traffic

Consistent campaign codes so we knew which channels and programs delivered net new names, pipeline and revenue

Optimization over time to get more and more efficient

Predictive analytics to score leads better and a nurture approach
To ensure we were passing quality leads to Sales, we needed to layer on the ability to score and nurture all leads.

Some of this we could do in Marketo, but to do a really good job, we needed some predictive analytics. For that, we turned to Lattice Engines – a predictive scoring platform that has really paid dividends.

What Lattice does is to analyze people/companies who actually bought our products, looking at many hundreds of data points for

↓

A word on campaign codes
The religious and consistent use of campaign codes for all marketing activities was absolutely critical to our marketing operations success.

In the past, we'd let the campaign code hygiene slip a bit and our analytics suffered badly. Now we have a locked-down campaign code protocol and if anyone breaks it, we take them out back and.... Tough but fair.

We'll discuss this more in a later chapter.

each one – everything from job title, company data including credit ratings, hiring profile, technology profiles and location to behavioral data, and shoe size (kidding, but it's only a slight exaggeration).

Lattice then looks at all leads to score them (and continuously re-score based on any additional response data) according to how similar they are to people who actually went on to buy. Leads are scored as A, B, C and D, with A being the highest propensity to buy.

In practice, the leads that scored 'A' turned out to be on average 6 times more likely to buy than the average leads we gave to Sales in the old days.

When that happens, the dynamic completely changes: Sales starts looking forward to leads from Marketing – to trust the marketing machine.

Our Lattice models get better and better the more we use it. We can look at different account profiles and coverage models and retrain the algorithm as we make changes to our marketing and sales methodology to keep improving. Hot stuff.

→ **What our predictive analytics does for us**

High-quality, high-fidelity lead scoring – trained on conversion to revenue (not just pipeline)

Far higher conversion rates

Boost of sales productivity, because we are no longer asking for follow up on D leads (they go into nurture)

A major credibility boost with Sales

Faster revenue!

Data integration and governance

The fifth piece of our marketing operations foundation was the ability to connect up all our data in a reliable way. This includes data integration between the various applications – but it also means taking all steps necessary to be able to join data from different systems down the road.

For example: We made sure that for every visitor, their Marketo munchkin (cookie) was communicated to our Adobe Analytics system, using Marketo API integration with Adobe Analytics. So we could have full confidence that we could directly tie our Adobe Analytics data to each prospect and their sessions.

A key part of all of this – and the entire big data marketing operations journey – is data hygiene and governance. To get the most out of big data, data quality is just as important as it is for traditional data warehouses.

When your entire revenue machine is built on data, you need to know the data is clean and accurate. The rigorous campaign codes discussed above (and intelligent campaign naming) are a great example of that, but it is only one of the new processes and policies we needed to implement in order to keep our data quality high.

If you take nothing else from this entire book, we hope you take this: **without consistent data rules and tags, you cannot do the analysis and tracking that big data marketing depends on. It's the only way to connect the dots all the way to revenue!**

And because we're in B2B, data mastering and hierarchies are another important component of governance: we're using our own Master Data Management platform so that leads and contacts are de-duplicated and all relationships are perfectly documented, e.g. which site a contact belongs to, which company the site rolls up to (we use data from our partner Dun and Bradstreet – everyone knows DUNS numbers, right?).

With data governance and data integration you can ensure you've got unique campaign codes and that they're attached to the

opportunity created in Salesforce after you pass the lead over (the essential piece of the attribution puzzle).

Without it, the world goes dark.

→ **What our data governance does for us**

Keeps our data clean, accurate, and trustworthy

Lets us track all the way to conversion (campaign codes!)

Establish the first and last touch and everything in-between for every sale

Whew.

So there you have it: our five big marketing operations foundations.

We had all this in place before the 60-day sprint that we talked about in the introduction – so now you can see that our starting line was already well down the track.

There's no short cut to this kind of thing. The engine is only as good as each moving part and – above all – the quality of the data you pour in.

If you're tempted to dive into a big data marketing program or to start a Data Lake tomorrow – worrying about the 'boring stuff' later: sit down and reconsider.

“Marketing has an opportunity, with technology, to capture the voice of the customer and help the rest of the company tap into what makes buyers tick.”
Laura Ramos, Vice President and Principal Analyst, Forrester

The big data marketing technology stack

We often hear marketers discussing their operations and saying, "It's not about the technology."

Well, big data marketing is about the technology. Of course, it's also about a mindset, an approach, and a strategy, but there's no way you can harness the power of big data without the right tech.

And, with the splintered state of marketing technology today, that means putting together your own marketing technology stack.

So this chapter is all about our technology stack and how it might influence yours.

Integrating your own technology stack may sound scary. It needn't be. We did it with a small team of mostly not-so-techie people, supported, when needed, by our awesome colleagues in IT.

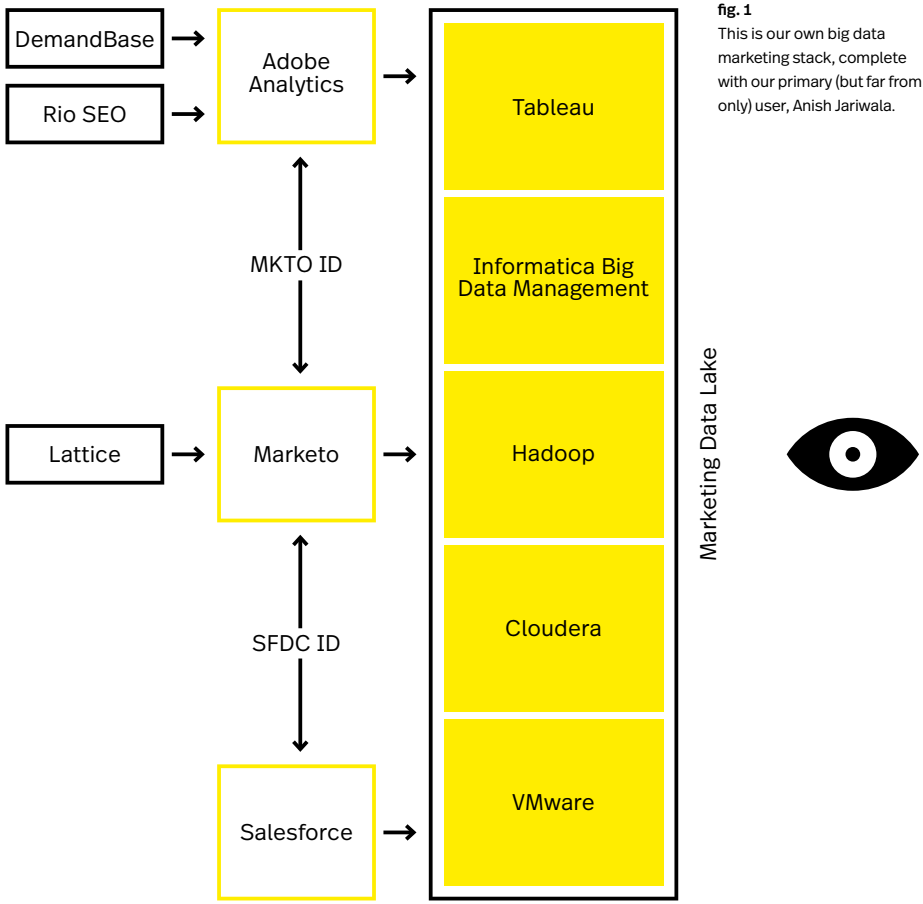
So let's do this.

Our big data marketing technology stack

As we take you through our technology choices, please remember that they're only that: the choices we made based on our specific situation, goals and needs.

You will no doubt make different choices. But you can still integrate your stack around your own Data Lake.

So here's our stack:



↓
The Stackies
Scott Brinker of Chiefmartec runs an annual award called the Stackies, celebrating the best tech stack visualizations. We love that idea and have learned a lot from the many visualizations submitted. Google it.

↓
fig. 1
This is our own big data marketing stack, complete with our primary (but far from only) user, Anish Jariwala.

The three pillars

There are three core applications that we'd expect to see in any B2B marketing department: CRM, web analytics, and marketing automation.

Together, they streamline, automate, and track the core revenue generation processes. It's hard to picture an effective B2B revenue machine without all three. Here are the three we chose:

1. Adobe Analytics
You need web analytics to track all visitors, especially the still-anonymous ones (the bulk of most B2B traffic).

We chose Adobe Analytics (formerly SiteCatalyst) because it's feature-rich, plus it works well with our Adobe Experience Manager web publishing environment. Its strength in things like web personalization, digital asset management, metadata capabilities, and tag management made us value an analytics package that was tightly integrated.

Adobe Analytics is also part of an active ecosystem, so there are lots of developers adding to it all the time. That's important to us. While there are a lot of amazing analytical tools out there, we don't want to be stranded on a platform that isn't constantly evolving.

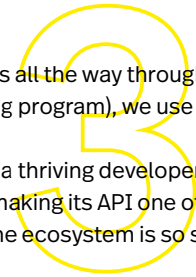
We find it's pretty easy to set up new reports and dashboards in Adobe Analytics, even for non-technical users, and we do lots of reports and dashboards.

2. Marketo Marketing Automation
For tracking known visitors, email and automating our nurture flows, we chose Marketo.

As mentioned in the Foundations chapter, we migrated from Eloqua to Marketo, largely for ease of use reasons. Is it perfect? No. We wish its APIs were a lot more developed than they are today and trust that they will develop over time to ease integration.

But Marketo is a rich, powerful marketing environment and we're making the most of that.

3. Salesforce CRM
For tracking the sales opportunities all the way through to revenue (a key part of any big data marketing program), we use Salesforce.



Salesforce is a great platform with a thriving developer/partner ecosystem. It also invests a lot in making its API one of the best in the business (one of the reasons the ecosystem is so strong) and that makes our lives easier.

In a way, these three pillar applications also create the need for the rest of our marketing operations stack. Because, as powerful as they are, they just don't solve all of the marketing challenges.

In fact, because they're separate applications, they're actually part of the data fragmentation problem. And that's where the rest of the stack comes in.

Big data management

Big data marketing is all about treating data as a strategic business asset. But a lot of marketing departments get seduced by the eye candy, jumping straight into the visualization tools and dashboards.

That's not only superficial, it's downright dangerous. Because unless your data is in order – cleaned, mastered, tagged, and secured – those pretty dashboards will be hiding all sorts of nightmares.

Here are the core data architecture parts of our stack:

Dynamic Tag Management
The heart of our data layer is Dynamic Tag Management (DTM), an incredibly powerful (and free) tool that comes with Adobe Experience Manager and integrates with our Adobe Analytics.

DTM is essentially like a CMS for JavaScript, tags, and tracking codes. It lets mortal marketers like us manage the campaign activations and data collection for our many marketing apps (not just the Adobe ones) – and do it from one, central web dashboard.

We start with the business logic – what we're trying to do – then use DTM to determine when to fire a tag, what data to collect and where to send it. It's all rules-based and driven by the events and conditions we choose. But because it's a single tag manager for all of our marketing apps (from Marketo and Adobe Analytics to Rio SEO, Demandbase, D&B, and Lattice Engines), it streamlines and automates what would otherwise be a hugely repetitive, manual process.

The Data Lake itself

We hold all the data from Salesforce, Marketo, and Adobe Analytics in a Hadoop cluster with 5-7 nodes: our Data Lake.

Suffice it to say that all our data lives here, in its raw, unstructured (or minimally structured) state, ready to be analyzed.

It's common to see big data marketing teams host their Hadoop cluster in a public cloud environment like Amazon Web Services or Microsoft Azure (and many of Informatica's customers do that). But we host our cluster on VMware machines in our own data center.

The right choice for you will depend on your current in-house infrastructure, the expertise of your IT teams, and the cost implications. For us, it was more economical to use our shared internal infrastructure – especially as we aren't dealing with big spikes in volumes or loads (so we don't really need the elasticity that the public cloud brings).

One day, we may move to the cloud, but for now, the VMware deployment works well.

Our enterprise data warehouse

Though this is a Data Lake program, we did use our existing data warehouse to get us up to speed.

At the beginning, instead of taking our Marketo and Salesforce data directly into our Lake, we took it from an operational data store that is used to feed the data warehouse. It wasn't ideal as there was some subsetting of data happening for the data store (although what we load into the warehouse is a much smaller portion of data) but for speed of deployment, it was the easy choice and gave us 95 percent of what we need.

In a perfect world, we'd feed this data directly into the Lake. But you probably don't live in a perfect world either, and it might make sense for you to use an existing warehouse to bridge over to your Lake.

Informatica data management tools
We use our own Informatica data management products to streamline and automate data management and ensure data quality.

This book isn't a sales pitch, so we won't bury you in features and benefits. To find out more, visit the Informatica big data management solution pages. In short, here's what we are relying on:

- Data Integration – for combining data from many different sources (you may have heard about PowerCenter, the data integration platform. We're also using our Cloud Integration product – it's easy to use and get started with).

- Data Quality – supporting our data governance policies in a repeatable, automated way.

- REV – our free data prep tool (which also has a premium, paid-for upgrade).

- Big Data Relationship Management – for exposing relationships in the Data Lake.

- MDM – mastering the most important domains (like accounts, products and contacts) and maintaining the hierarchy of contacts>sites>accounts (with help from D&B).

Knowing these products intimately is an unfair advantage for our team. But the tools are easy to understand and to start using, so you're not a million miles behind us here. And because of the popularity of the tools, there are tons of people with the right skills out there and they are easier to find than some of the specialized big data skill sets.

As you'd expect, we're huge believers in using tools to automate and accelerate the otherwise mind-numbing data preparation tasks that make for a smooth big data program. We won't pretend there aren't lots of other great tools out there – there are – but our tech stack picture wouldn't be complete if we hid the fact that we eat our own dog food.

Enrichment

We don't just rely on the data we can collect by ourselves (first-party data). We also enrich that data with third-party partners:

Demandbase
We use Demandbase to do reverse IP-lookup, allowing us to see what companies are visiting our site, the company size, and industry segments. That's hugely useful for Account Based Marketing, of course, but it also has some customer experience benefits on the site. For instance, we can use the Demandbase data to pre-fill forms or to personalize the site based on industry.

The Demandbase data comes straight into Adobe Analytics, or into Marketo (through forms), then on into the Data Lake.

Dun & Bradstreet
We enrich and validate our data with firmographics from our partner Dun & Bradstreet – the universally recognized Data Universal Numbering System (DUNS) standard. It helps that this integrates natively with our Master Data Management.

Rio SEO

Rio SEO specializes in local search marketing automation but we use it a bit differently: to identify buying teams within accounts and to track word-of-mouth influencers.

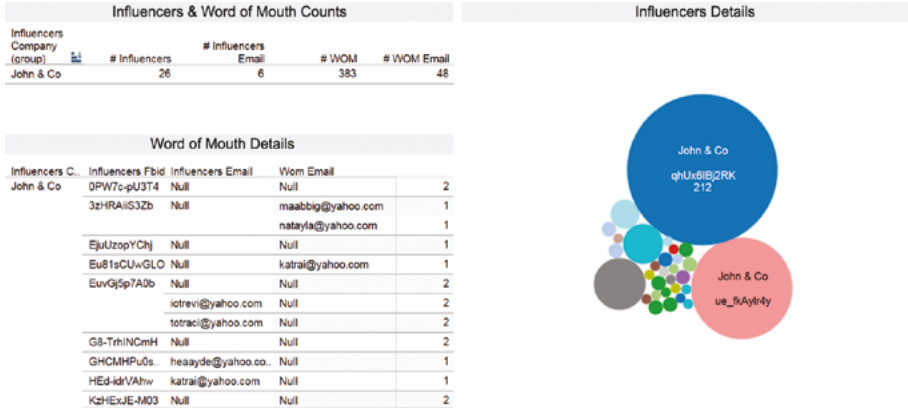
More than 5 percent of our site traffic comes via word-of-mouth. It's disproportionately important to us because it's among the most highly engaged traffic and because some of it indicates buying group relationships.

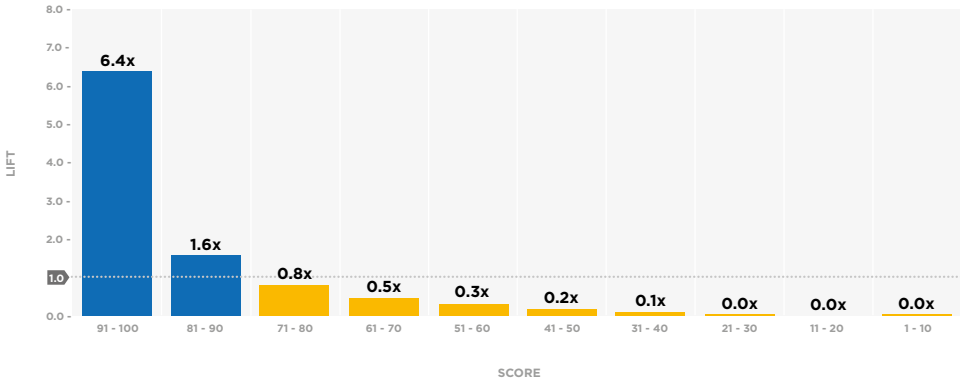
Here's how it works: When Sarah is browsing the site, each page URL has a Rio code. If she copies the URL to paste into an email to Bob or shares it in the social media, that code goes with the link. When Bob (or a person who clicks a link in social media) comes to the site and is known to us in Marketo, we can see that Sarah was the influencer who brought him to us.

We created 'bubble reports' to show us our most important influencers and the size represents WOM count they brought to us (fig. 2).

Clearly, this is gold dust for identifying buying teams within target accounts and spotting opportunities we didn't know about.

↓
fig. 2
Tableau dashboard: Influencers & WOMs relationships.





↓
fig. 3
Lattice Engines: Likelihood to convert by Lattice Score.

Predictive Analytics

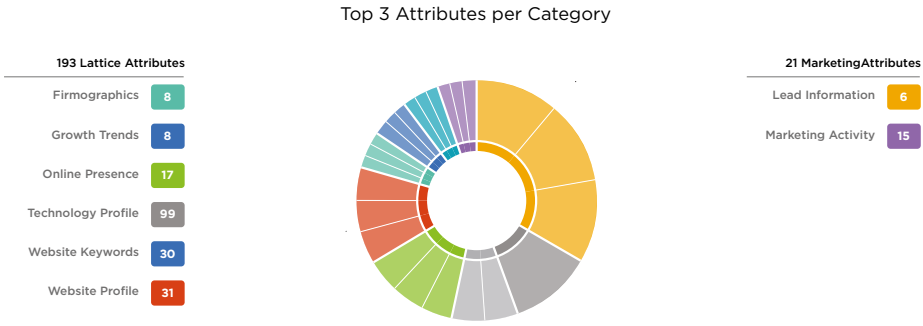
We use predictive analytics to score all prospects according to how likely they are to buy. The 'A' scores get special attention from our SDRs – which they love because an 'A' prospect is six times more likely to convert than an average prospect (fig. 3).

A, B, C scores are in blue. The average conversion rate is normalized to 1x, with 6.4x being the lift in the highest score group. All D leads (below average) are in yellow, the number scores are transformed into simple A, B, C and D buckets (aside from being easier to use than numeric ranges, it allows us to shift the score ranges based on model performance or changes).

Lattice Engines

Every predictive model has hundreds of dimensions that are either positive or negative predictors. To develop the predictive model (fig. 4) we analyzed more than 70,000 leads that met two criteria: we had company attributes for them and they actually went on to buy from us.

We tested the model against a control group of almost 20,000 additional leads to make sure we had a converged model. Now in production, the model analyzes our prospect universe for lookalikes and continuously scores and rescores all leads.



↓
fig. 4
Lattice Engines: One of our models with detail on the lift.

Right now, we have two models running: one for buyers of our licensed products and one for buyers of our cloud products (with slightly different models for North America and Rest of World so there are actually four in total). Each lead gets two scores depending on how likely they are to buy each kind of product.

You might think we'd have a model for each and every product but that would not only be overkill, there simply wouldn't be enough data to train each model accurately.

We work very closely with Lattice to keep verifying and improving our models, training them with opportunities that actually convert. That's important. Predictive analytics is generally only as good as your efforts. If sales or marketing practices and processes change, it's easy for a model to start drifting. As the models are trained on the past, you have to keep an eye on the future.

Big data visualization

The dashboard is what our marketers see when they practice big data marketing – and it's what most marketing teams spend the most time thinking about.

To us, data visualization is the 'last mile' of big data marketing, not the journey. You can use any visualization or BI tool you like as long

as your data management is robust. As a company, we work with pretty much every BI and visualization tool, so it's important we're Switzerland here. They're all really good!

Tableau
Tableau gives a lot of power to the non-technical user, allowing us to explore the data and test our hunches. And we were already familiar with it, so it was an easy choice.

Essentially, Tableau makes it easy to visualize data from different analytical stores including Hadoop in one pretty dashboard. We'll talk about this in the Dashboard chapter, so we won't drill down here.

Your own stack
So that's the tech stack behind our own marketing operation.

What your own stack looks like will depend on many factors, chiefly your budget, your team, your data, your legacy infrastructure, and your relationship with the IT department (ours is excellent).

But we imagine you'll need many of the pieces discussed above if you really want to drive your revenue engine, instead of letting it drive you.


Q
&
A

Q&A with Scott Brinker

Scott Brinker is the founder and editor of chiefmartec.com, the leading blog and resource site on the intersection of marketing and technology. He's also the creator of the annual Marketing Technology Landscape supergraphic, author of the book Hacking Marketing: Agile Practices to Make Marketing Smarter, Faster and More Innovative, CTO of ion interactive and program chair of the MarTech event series.

Finally, he's one of the smartest and nicest people in our field, so we're thrilled we were able to pin him down and ask him some questions:

Q **Marketers always seem to think they're one app away from a breakthrough. But we don't hear much about the hard work of bringing all those apps together. Is that what you're seeing?**

Scott Couldn't agree more. Marketing tech is in a fascinating place and there are hundreds of enabling apps, tools and platforms. But turning that potential into real marketing capability is about process and people

and organization – there's not enough discussion about that in the MarTech space.

Q The most recent Marketing Technology Landscape supergraphic shows a dizzying galaxy of alternatives for marketers. All this choice can be paralyzing. Is it viable to sit this out for a while until the landscape simplifies?

Scott It is a chaotic time. But even if you believe the idea that a dramatic vendor consolidation is coming (which I'm not sure I do), the question is still, "When?" Are you willing to sit out the next 2-3 years? Where will your company and your career be if you do that?

Empirically, this is the market you have to deal with today and there's a job to be done. You can let yourself get stalled by all the possible tools out there or you can take a "satisficing" approach to finding the tools that will let you move forward. In the end, it's less about the specific tool and more about how you use it.

Q Are there certain technologies that every B2B marketing team should have?

Scott Every company has its own circumstances but, as I touch on in Hacking Marketing, I think about marketing technologies in two groups: the core and the edge.

You need to standardize on a core set of systems that underpin your marketing operations. For most B2B teams, this will mean CRM, marketing automation, some kind of web experience management and analytics. For what you guys are doing, the Data Lake is clearly a key part of your core systems too.

Then there's the edge layer, where you can benefit from a different approach that is less about standardization and more about innovation. You can explore and experiment with different tools and techniques – new ways to engage

and connect with audiences. You don't need to be as standardized or controlled with these, when you're exploring them in the edge.

The key is to not try to manage the core and the edge in the same manner. You don't want to constantly re-engineer your core systems or you'll always be spinning your wheels. And you don't want to be so strict with edge tools that you don't provide room for play, learning and experimentation.

Q What's behind the rise of the Marketing Operations role?

Scott I think two factors are driving it. First, as the complexity of marketing explodes – so many more channels, campaigns and programs to manage – the operational components become more important to ensure you can connect and orchestrate all these moving parts.

Second, the explosion of data available to marketing and the changing nature of that data calls for a stronger operations role. Data used to be about high-level reporting of a relatively small set of limited-value data. Today, the data is the ground truth of the customer relationship. There's such an incredible amount of depth and insight in there if you have the capability to plumb it.

Being able to manage, massage and manipulate all this data to uncover insight and generate advantage is a natural role for the modern marketing operations professional. Today, a strong ops group is a major competitive win.

Q Will there come a time when marketing departments will get all of their marketing tech from one vendor?

Scott I don't think so. Among other factors, there's the problem of market power. Do you want one vendor to be able to control so much of what you do? We saw some of that captivity in the early days of things like ERP, and I don't think marketers are eager to repeat it.

One thing I like about the Data Lake approach is that it gives you control of your own data and your own destiny. That's a tremendous operational benefit even before you factor in the agility dividend.

Q Moving to agility then, you're a proponent of agile marketing management. Why is agility such an important dimension for marketers today?

Scott I think of there being two kinds of agility: Agility with a capital A and agility with a lower-case a.

The 'big-A' Agility is all about formal methodologies like Scrum and Kanban. I'm not religious about any particular method. There are strengths in all of them.

'Small a' agility is all about your ability to achieve outcomes. To deliver value to the business in a reasonable time frame. To be able to respond to changes and opportunities quickly. The problem is that everyone can claim this kind of agility. In the absence of a methodology, it's just one of those warm-and-fuzzy – emphasis on fuzzy – words.

That's fine, but I'd ask a CMO, "What specific things are you doing differently today that make you more agile?" My book is all about a series of mechanisms for achieving agility. And the Data Lake idea that you're pioneering is an example of that: an actual functional capability that makes your operation far more agile.

Q Some marketers feel that they can rise to the data integration challenge with a DMP. Is that realistic?

Scott So DMPs started in the ad tech world and I have seen marketers looking to use them to integrate all of their marketing data – internal and external. It's the idea of a Data Hub and that's not far from the Data Lake concept.

But whether DMPs can evolve from the ad tech space to play this role is debatable. Another approach from a different direction is the Customer Data Platform (CDP), a centralized system of trusted customer data that starts in the core and evolves out to handle things like ad tech or personalization or whatever.

At some point, it's a debate about labels. There are so many incentives to bringing all your customer data together in one place that it's almost inevitable.

For me, a DMP is about deriving a single identity for your prospects – and it's optimized for doing it at ad-tech speeds, in milliseconds. But identity is only a very small piece of the overall data picture. No DMP in the world comes close to the richness and agility of the Data Lake. But a DMP as an identity server that feeds into a Data Lake? That might be a good way to handle it.

“Under conditions of complexity, not only are checklists a help, they are required for success.”
Atul Gawande, 'The Checklist Manifesto'

Two killer checklists

By now, you're seeing that your big data marketing program must be supported by consistent and high quality data.

A big part of ensuring consistent data is to create data governance processes and policies and then make sure everyone follows them religiously. The other big part is correctly setting up your CRM, marketing automation, and analytics to capture and analyze all that relevant data (fields!) and report it correctly (filters!).

When we think about making complex processes simpler, we think checklists (a favorite book is Atul Gawande's The Checklist Manifesto, all about how simple checklists can make dramatic impacts in complex systems).

- **It's okay to skip this...**
  If you're not working regularly in a marketing automation or an analytics suite, this chapter may be a bit... opaque. So just skim it to get an idea of the kinds of issues you'll hit when you set these up – or skip it altogether. We won't be offended.

Our checklists

Since we use Adobe Analytics and Marketo, our two most important checklists are for those two tools. If you use these applications, you may want to use these checklists too, adapting them for your own processes.

But even if you don't use either of these tools, the checklists will still give you a good sense of the kinds of set-up and governance policies you need for accurate big data marketing.

Fortunately, Anish has a wealth of experience with both Adobe Analytics and Marketo. So he drew up these two checklists to keep us all on track – and keep our data healthy. We've annotated each list to help explain the thinking behind them.

The Adobe Analytics Checklist

Setting up your analytics properly is absolutely essential for big data marketing. This checklist helps us keep on track:

- ☐ **1. Create one Global Reporting Suite** for all your relevant web properties.
- ☐ **2. Think through exhaustive business questions** that you want to answer for all your web properties. Based on this, assign sProps, eVars, and success events optimally.
- ☐ **3. Use friendly page naming conventions** for every web page.
- ☐ **4. Create Product and Solution eVars and success events.** [This is essential if you're a multi-product company.]
- ☐ **5. Tag all the Marketing Automation (e.g. Marketo) landing pages in Adobe Analytics.** Make your CMS and Marketo landing pages seamless from a reporting perspective. [Form Submits will be your most critical event.]
- ☐ **6. Deploy Marketo SOAP/REST and Demandbase APIs** on the website.
- ☐ **7. Set up Marketing Channels** for web traffic attribution.
- ☐ **8. Assign intelligent external campaign parameters** to track paid traffic.
- ☐ **9. Remove internal (employees/ partners) and robot traffic** from the Reporting Suite.
- ☐ **10. Avoid unnecessary bells and whistles** in your analytics implementation – anything there's no customer for.

Notes for the Adobe Analytics Checklist

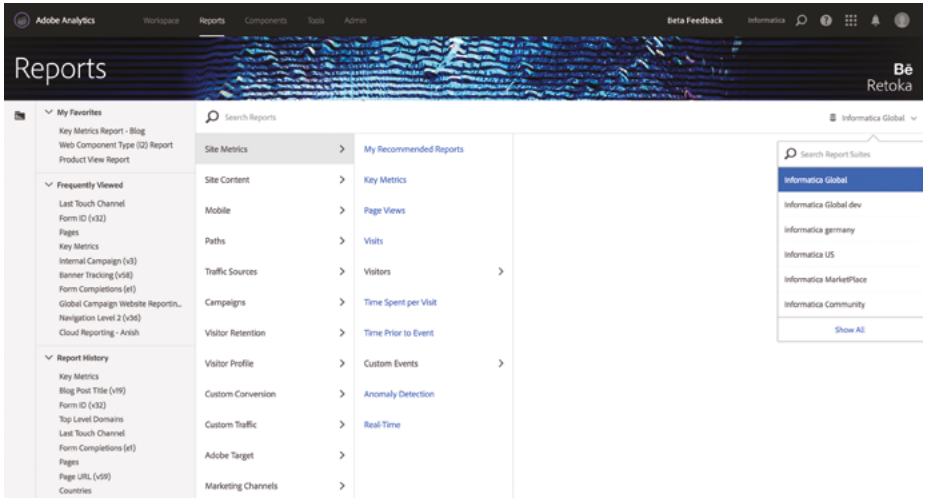
1. **Create one Global Reporting Suite for all your relevant web properties.**

→ This is such a simple concept, but it's overlooked in many Adobe Analytics implementations that we've seen.

All your pertinent web property traffic should consolidate into one reporting module in Adobe Analytics. If you don't do this you will double count your visitors and you won't be able to frame their end-to-end web journey.

For example: If your main site traffic goes into one suite and your blog site goes into another, then a visitor who navigates to the blog portal from the main site will be counted as an individual visitor in both suites. Instead, the visitor should be counted as a single visitor and you should be able to see in Analytics that the person went to the blog site from the main site. This is critically important.

↓
fig. 5
Adobe Analytics: Set up your Global Reporting Suite.



2. **Think through exhaustive business questions that you want to answer for all your web properties.**

→ Based on this, assign sProps, eVars, and success events optimally.

In any analytical framework, you have to think about two things: dimensions and metrics. You can break down metrics by different dimensions to derive analytical insights.

In Adobe lingo, eVars are dimensions, Success Events are metrics. sProps are similar to eVars but they're not persistent, changing value as the visitor goes from page to page – they're mainly used in pathing analysis. (BTW, we sincerely hope Adobe streamlines eVars and sProps and just comes up with one kind of dimension.)

Setting up your sProps, eVars and Success Events should never happen in isolation from the business. It's absolutely critical that your analytics reflect your business goals and the metrics that track them.

Get it right and your Adobe Analytics will make it easy to do 'mass customization' specific to your business and your conversion tracking (whether it's revenue, orders or form fills). But keep in mind: as your website changes, you need to be mindful about the impact on your Adobe Analytics implementation. Adobe Analytics is a living, breathing, listening system that watches your visitors navigate around your website. If it's not set up to align with your business, you'll only get half the value.

3. **Use friendly page naming conventions.**

→ One of the cool things about Adobe Analytics is that you can provide a friendly name to each and every page that is rich, descriptive and readable (www:us:en:products:products-portfolio) or use super-simple URLs (www.informatica.com/products). URL names can get very complex very quickly, so set up a solid page naming convention and stick to it.

4. Create Product and Solution eVars and success events.

→ This is important for any multi-product company.

We use the mass customization in Adobe Analytics to give us a product-centric view, similar to that of any ecommerce business.

As you go to any retailer website and navigate to, say, a vacuum cleaner, Adobe Analytics can capture the product name (brand/make/model), the category, and how many vacuum-cleaner-related pages you saw. Since we're a multi-product company with a diverse product portfolio, we set it up the same way for our website. Knowing what products and solutions our prospects and customers are interested in is gold dust.

↓
fig. 6
Adobe Analytics: Custom Conversion → Product name (eVar6) and Success Event (e10).

Enter term to filter data | Go | Advanced | Metrics

	Product Name (v6)	Product Views - Custom (e10)	
1.	data quality|data quality services|addressdoctor	51,606	21.0%
2.	data integration	35,207	14.3%
3.	cloud integration	28,459	11.6%
4.	big data	19,930	8.1%
5.	master data management	19,479	7.9%
6.	data quality	12,684	5.2%
7.	data integration|powercenter	9,198	3.7%
8.	cloud integration|cloud connectors	6,677	2.7%
9.	data quality|data as a service	6,161	2.5%
10.	integration platform as a service	4,515	1.8%
11.	cloud	3,726	1.5%
12.	data as a service (daas)	2,920	1.2%
13.	data security|data masking	2,913	1.2%
14.	master data management|product information management	2,436	1.0%
15.	data integration|real-time integration	2,251	0.9%
16.	big data|big data edition	2,146	0.9%

5. **Tag all the Marketing Automation landing pages in Adobe Analytics.**

→ It's important to make your CMS and Marketo landing pages seamless from a reporting perspective. That's especially important for Form Submits, the most critical Adobe Analytics event.

This can be a big undertaking since Marketo and Adobe Analytics are very different tools. When you visit informatica.com, it's hosted on our CMS, Adobe Experience Manager. Our landing pages (with web forms) are hosted in Marketo, so when you navigate to our landing pages we need to tag them in Adobe Analytics. That allows us to track visitor behavior seamlessly across CMS and Marketo pages.

When we set this up, we had more than a thousand landing pages globally for paid media, gated content assets on the website, etc.. They were all built in Marketo and needed to be tagged in Adobe Analytics. Not a trivial task but well worth it (and Tag Manager is a godsend here).

6. **Deploy Marketo SOAP/REST and Demandbase APIs on the website.**

→ Another critical element to map end-to-end customer journeys between Adobe Analytics and Marketo is to have a common ID shared by these two systems. This common ID allows us to stitch Analytics and Marketo data together to solve some of our most important use cases.

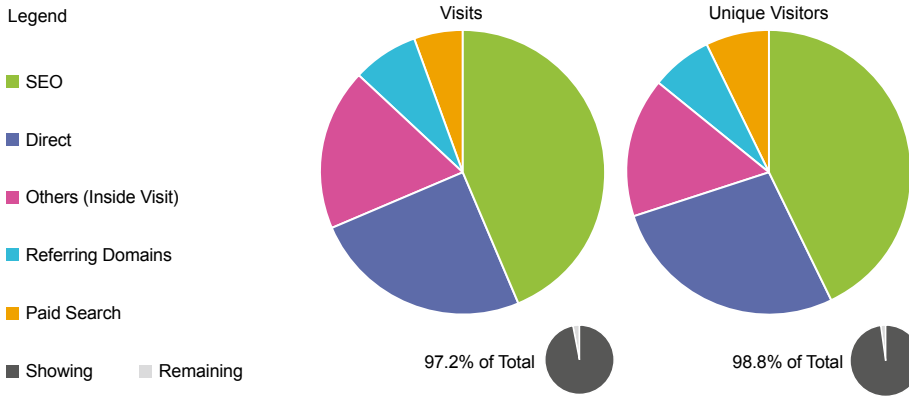
For us, this common ID is called "Marketo ID" and it's unique for every visitor. We expose this ID to Adobe Analytics using Marketo's SOAP and REST APIs so we can track anonymous visitors becoming known visitors in the Data Lake.

We do the same for integrating our Demandbase data, which gives us the reverse IP lookup that lets us enhance our profiles with company and sector information.

7. Set up Marketing Channels for web traffic attribution.

→ Marketing Channels is an out-of-the-box feature in Adobe Analytics and it's easy to set up. It allows us to track Paid, SEO, Direct and Referrer traffic separately so we can measure engagement and conversions across these channels. Invaluable for budget optimization.

↓
fig. 7
Adobe Analytics: Marketing Channels → Last Touch Channels.



Last Touch Channel Report | All Visits (No Segment) | April 2016 | Graph generated by Adobe Analytics at 4:46 PM PDT, 21 Apr 2016

8. Assign intelligent external campaign parameters to track paid traffic.

→ All external paid traffic URLs should be tagged with consistent parameters so that you can track the paid traffic that's coming to your website.

For us, the paid traffic is Display, Remarketing, Paid Social, and Search. The naming convention for parameters needs to be smart enough so that we can slice and dice this traffic by regions, products, marketing themes and channel. This can involve a lot of manual hard work – for us, it meant tagging those 1000+ URLs that are in production.

9. **Remove internal (employees/partners) and robot traffic from the Reporting Suite.**

→ It's important to remove internal (from the Informatica IP address) and any partner-related traffic (agencies, etc.) from your reporting. Otherwise, this traffic will significantly distort your analytics. Also, watch out for any robot traffic hitting your site on a monthly basis and remove that from the suite.

10. **Avoid unnecessary bells and whistles in your analytics implementation – anything there's no customer for.**

→ We've often seen people put unnecessary functionality into their analytics that no one cares about. This isn't just a waste of set-up time, it also becomes a burden as the website generates all sorts of maintenance issues.

In our case, we were coming from the opposite end: the old SiteCatalyst instance was plain vanilla and we weren't coming close to leveraging all that it had to offer. So the temptation is to go 'all-singing, all-dancing' from the start.

It takes discipline to focus on business needs and to prioritize your features and functions into a sensible roadmap. Do it up-front and you spare yourself needless pain.

Those are the notes for our Adobe Analytics Checklist.

Now for the marketing automation side of things.

The Marketo Data Governance Checklist

We use Marketo to track known web visitors and execute and monitor campaigns and ongoing content marketing nurture flows.

If you're deploying Marketo, run through this list and see how you're doing. Skipping any of these could mean you're storing up problems for later.

- ☐ 1. Do all paid programs have period costs?
- ☐ 2. Do all relevant programs have the correct Channel and Success statuses defined?
- ☐ 3. Have you set up exhaustive program tags?
- ☐ 4. Have you set up your website-related programs?
- ☐ 5. Are contact roles attached to every opportunity?
- ☐ 6. Check the correct set up of Opportunity object, Opportunity amount, and Expected Revenue
- ☐ 7. Is "Acquisition Program" set up for all leads?
- ☐ 8. Is a well-negotiated revenue model up and running? Will you be using multiple models?
- ☐ 9. Is there a Group Field Organizer for the Revenue Model set up?
- ☐ 10. Are all relevant custom fields imported into Revenue Cycle Analytics in Marketo?

Notes for the Marketo Checklist

1. **Do all paid programs have period costs?**

→ We run many paid marketing programs such as PPC, offline events, social advertising and retargeting. It's important to capture the total cost of each marketing activity (a rough approximation is fine) so you can do ROI calculations: the revenue contribution (by particular program or channel) divided by costs.

We've seen people in meetings say, "Look: that content got 200 downloads, let's do more like it!" without even knowing how much was spent to get those downloads. That 'winner' could be a loser in terms of real ROI.

In truth, Marketo makes it very difficult to manage the cost side of campaigns. So we confess that at times, we do our ROI calculations outside of our analytics environment, exporting data into a spreadsheet for easier analysis. We hope this will change soon.

2. **Do all relevant programs (campaigns) have the correct Channel and Success statuses defined?**

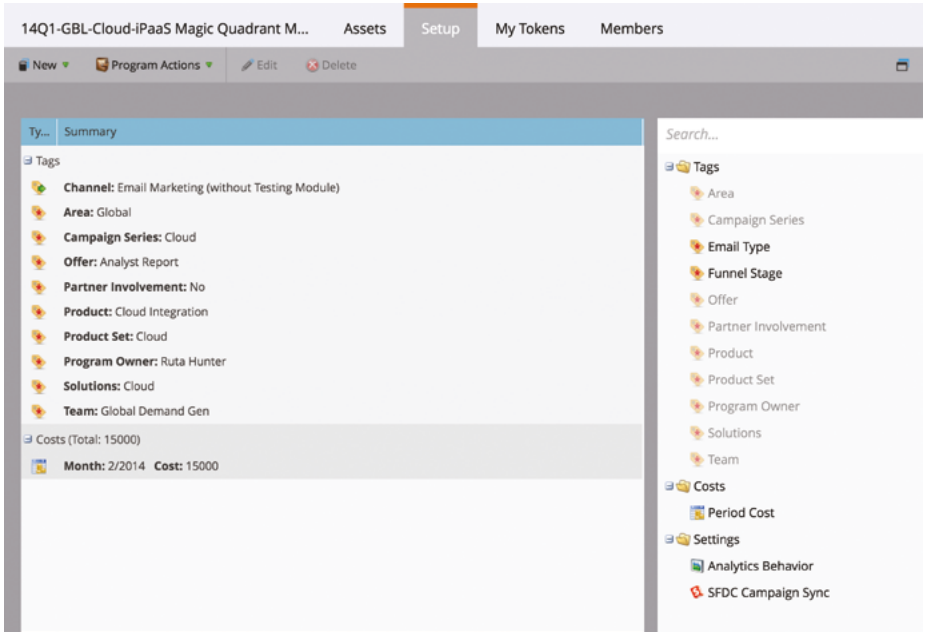
→ Every program in Marketo has to be assigned to the right channel and have the right success statuses. For example, for an email program, success might be clicking a link in an email. For tradeshows, it could be attending a specific session or visiting our booth. If channels and success statuses are not sorted out before you go live, than you can't capture the whole value of your programs in Marketo.

Do this up front because it's very hard to fix these issues once you've gone live.

3. **Have you set up exhaustive program tags?**

→ Tags are the easiest way to flag all of your programs in analytics, across various dimensions such as regions, themes, marketers, etc.. Once you have tags, you can slice and dice the analytics for programs by these dimensions.

IF YOU DON'T HAVE THE TAGS, YOU CAN'T GET THE INSIGHT.



↓
fig. 8
A Program setup in Marketo with tags, period costs, etc.

4. Have you set up your website-related programs?

→ Between 60 and 80 percent of New Name acquisitions happen on our website, including Contact Us, Pricing Pages and Product Demos.

Prospects who are aware of your company come to the website directly – it's 'free' traffic and it's often highly engaged. You need to have an exhaustive setup of Marketo programs on the parts of your site that drive lead generation. Many times, companies set up outbound programs in Marketo such as email, but ignore the inbound programs or channels, such as their own website.

5. Are contact roles attached to every opportunity?

→ It's very important that Opportunity Contact roles (that

is email addresses) are explicitly attached to every opportunity object in Saleforce.

If sales folks don't tie an email address to each opportunity, it's very hard to see which Marketo programs contributed to a given opportunity. To mitigate this problem, we use account-based attribution in Marketo – it's not precise but it helps. (We have some other solutions in the works to mitigate this challenge.)

Sales has to identify the buying teams and associate their names to the opportunity in Salesforce. If these names are not associated with the opportunity – the right hand side of the revenue cycle model (see 8-10 below) will not work. This is a big challenge in B2B Marketing. Almost 50 percent of opportunities at Informatica do no have contact roles associated with the opportunity. The other 50 percent of the opportunities have only one name called out explicitly. To mitigate this problem, we use an account-based view for attribution within Marketo. But that doesn't solve the problem in this model and the analytics would remain incomplete for the right hand side – the opportunity related stages. We're sure this is a common issue across a lot of companies. Some day, we'll convince our friends in Sales to help us out!

6. **Check the correct set up of Opportunity Object, Opportunity Amount, and Expected Revenue.**

→ Marketo allows opportunity fields to sync over only as a dimension and not as a metric or measure. It's very important to understand these fields and their implications for marketers.

For example: you need to know the currency used in the "Opportunity Amount" field in Salesforce. If you're a global company, regions could be using their local currency to populate this field. We use a converted currency field to map to Marketo's Opportunity Amount so we can do an aggregated analysis in one currency.

7. Is "Acquisition Program" set up for all leads?

→ Every lead in Marketo has to have an acquisition program. Acquisition credit is given to the Marketo program that helped in converting an anonymous lead to a known lead (typically using form submits). For a given lead, there will be only one Acquisition Program.

8. Is a well-negotiated revenue model up and running? Will you be using multiple models?

→ This is critical if you're going to use the Revenue Cycle Analytics (RCA) module in Marketo. We worked closely with Field Marketers, Sales Development Reps, and Sales to define our revenue model, capturing the journey of the anonymous lead through becoming known and all the way to the opportunity.

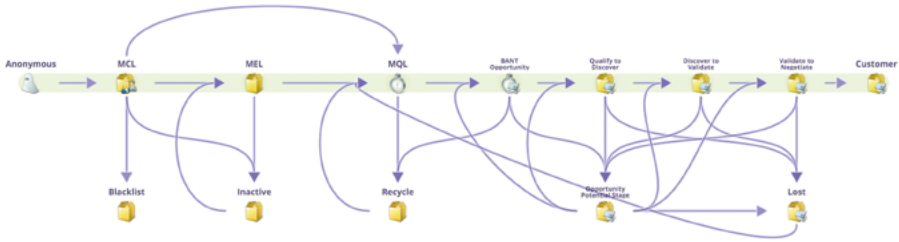
9. Is there a Group Field Organizer for the Revenue Model set up?

→ The Group Field Organizer helps in aggregating various values of a given field to use in Revenue Cycle Analytics. For example, Lead Score may have a value ranging from zero to infinity. You may want to create buckets of 0-50, 51-99, and 100+ scores so you can analyze leads this way.

10. Are all relevant custom fields imported into Revenue Cycle Analytics in Marketo?

→ Salesforce allows most of the leads and critical opportunity fields to sync into Marketo RCA for analysis in the form of dimensions and measures. Make sure you bring along the custom fields that will help you with your analysis.

↓
fig. 9
Revenue Cycle Model.



Make your own checklists

These two checklists capture a lot of the lessons we learned as we built our big data marketing machine. We hope they help accelerate your own initiatives and keep them on track.

Just as importantly, we hope you'll create your own checklists as your journey progresses. This 'tribal knowledge' is a valuable business asset – but only if it's captured and shared.

Now let's talk about data hygiene and data management – a hugely important part of any big data marketing program.

“It is a capital mistake to theorize before one has data.”
Sir Arthur Conan Doyle

Managing your data

The critical part of any Marketing Data Lake is the data itself. So this chapter drills down into our data.

For a lot of marketers, the word ‘data’ sounds clinical and mathematical – a bunch of characters and numbers, like those in a financial spreadsheet. To us, core data comes first and foremost from listening to your customers – seeing what interests them, noting what they respond to, and what they ignore – and that’s the single most important thing you can do as a marketer.

In short, big data marketing is a huge opportunity to look more closely at your customer experience, end to end:

– To segment your prospects intelligently

– To market to people efficiently and effectively

– To personalize each interaction and the entire customer journey

– To optimize your marketing budget and maximize its impact

All that is available to any marketer.
But only if you listen to your data.

The data challenge

Unfortunately, the data that you collect from your customers isn't pre-digested for you. It's a whirlpool of disparate data streams from a wide variety of marketing channels and sources and applications.

And that's the essence of the Data Lake challenge:

How do you get clean, complete, trustworthy customer data and associate it with accurate profiles – when it comes from multiple sources, under different names, email addresses, and devices, and is plagued by poor form fills, major data gaps, duplicates, and conflicts?

The challenge is tough enough that most marketers simply give up and accept a fragmented customer view as the price of doing marketing. That's a shame because solving the problem is well within reach. Here's how we did it:

The data we work with

Broadly, we collect two kinds of data:

Explicit data – The things people tell us by filling in our forms (name, address, job title, company, email, phone number… we use progressive profiling to enrich this data over time and not ask for too much in any single form).

Implicit data – The things we collect to complement and

enrich that explicit data (click streams, opens, pages visited, content viewed, etc., as well as purchased data like firmographics, technology profiles, hiring profiles, credit ratings, etc.).

And, as a reminder, we get our marketing data from a few, primary sources: web analytics, marketing automation, CRM and some from social media.

A lot of this data is accrued to the individual customer profile – but this is B2B, where most purchases are made by buying teams. So we also want to know about each prospect's relationship to other people, to their company, and to their site.

Understanding that hierarchy is important for B2B selling, but most lead-gen stacks are almost completely oriented around the individual lead. And even Account Based Marketing approaches typically lack the capabilities to identify buying teams, often lumping everyone at a company into the same bucket.

As you can see, the data we work with is exactly the same as the data that you probably work with: the big pillars of any B2B revenue generation infrastructure.

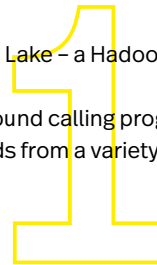
What makes us a bit different – and what makes this a big data play – is what we do with that data and how we manage it.

How we do it – in five steps

Marketing data management is a big topic, so we'll break it down into five steps: Collect, Clean & Validate, Enrich, Use, and Govern:

1. Collect
All of our data is brought into our Data Lake – a Hadoop cluster.

For example, if we're running an outbound calling program, we build the list by pulling relevant records from a variety of sources;

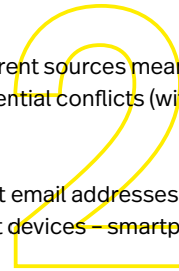


uploading it into Marketo; using predictive lead scoring for prioritization (Lattice Engines); pushing it into Salesforce as tasks for our Sales Development Reps; then recording the results – positive and negative – in Salesforce (via sync this goes back into Marketo as well).

For a trade show or other event, we load the data in as a list import (with a program name, code, and the cost of the event), or we might use iPads for on-site data capture, sending form fills straight into Marketo.

The motion is essentially the same: we time-stamp every interaction (or attempted interaction), associate it with a source and a program, and then link it to all relevant profiles.

2. Clean and validate



Combining so much data from different sources means there will be a lot of data duplication and potential conflicts (with slight variations in names, etc.).

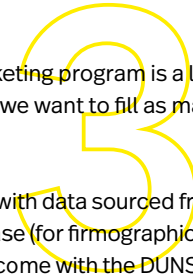
A person might use several different email addresses when they interact with us. Or several different devices – smartphone, tablet, different desktop browsers…

To make sense of it all, we use Informatica's own Master Data Management (MDM) suite, mastering two dimensions: customer and company. MDM (and our Data Quality toolset) helps us collapse all of that activity to one profile, an essential step for all future targeting and tracking.

MDM is an automated process guided by our own business matching rules. If the system sees two records for the same individual, it automatically collapses them together – as long as the confidence level is above the threshold we set. If it's not sure, it kicks up the exception to a data steward who can decide. If a lot of exceptions get resolved in the same way, the system learns that rule and gets smarter. You can set the sensitivity wherever you like – we prefer to keep profiles separate until we're sure they're the same person.

Mastering and cleaning the data is important, but you can never assume that the data you've collected is actually correct and usable: people make mistakes entering their addresses, give wrong phone numbers and email addresses, put the state and zip in one field… and the result can lead to a mess.

Again, we use Informatica's own Data-as-a-Service tools here (shameless plug, we know) for contact data validation. They make sure that every postal address, phone number, and email address is correct and in our desired format – BEFORE the data enters the database.



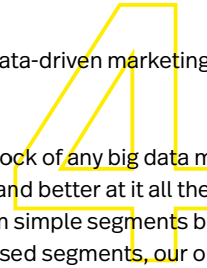
3. Enrich

Since the goal of any big data marketing program is a lot of clean, rich, and complete profiles, we want to fill as many gaps as possible.

To do that, we enrich our own data with data sourced from partners and suppliers, including Demandbase (for firmographics) and Dun & Bradstreet (for all the goodies that come with the DUNS number).

Adding third-party data to a big-data engine like ours is pretty simple: we load it in, match it to our records, merge and import the sets using our data integration platform and clean it with our data quality and validation tools.

Part of enriching is connecting each profile to people in the same company, at the same site (using MDM) or in a connected buying team (early days for us but we use Rio SEO to connect the word-of-mouth threads as people share our URLs with colleagues).



4. Use

Using our data means deploying data-driven marketing programs, usually to specific segments.

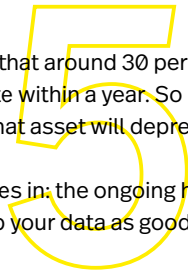
Segmentation is a core building block of any big data marketing program and we're getting better and better at it all the time. For instance, when we moved from simple segments based on job title or industry to behavior-based segments, our open rates

and click-through rates increased by 3-4x. It's all about relevance.

An essential behavioral segmentation strategy for us is to target by product interest. We track which web pages and content people are consuming and what emails they respond to, so we can capture product interest data in their profiles. That helps us target more accurately and increase engagement.

One of our early use cases is Account Based Marketing – marketing and selling to a named list of accounts. The work we do of connecting individuals to accounts and identify buying teams, as well as develop target segments for digital and calling campaigns, is an essential driver of this work. (More on ABM later.)

5. Govern



Data never stands still. Experts say that around 30 percent of any marketing database goes out of date within a year. So if you just clean and master your data once, that asset will depreciate fast.

That's where data governance comes in: the ongoing hygiene processes that make it easy to keep your data as good as it can be.

Part of that process for us is Data Profiling: assessing the current state of our data, and reporting on things like:

- Number of records and fields
- Duplicate percentages
- Fill rates for different fields
- Compliance to the definition for that field (so a name isn't made of numbers, for instance)

That gives you a nice snapshot of your data hygiene and flags any issues – so you don't segment based on a field with low fill rates, for example.

Another part of data governance is a clear set of policies backed by

good communication and training so that everyone who touches the data knows the rules and why they're important. The Checklist chapter is all about that.

Manage what matters

This quick overview of our approach to marketing data management should convince you of the importance of the data part of the whole Data Lake approach.

You can use all the fancy techniques in the world, but if your data isn't clean, complete, and trustworthy, you'll end up going down the wrong roads. The good news is that all this data management is easier than ever before, with so many tools available to automate and streamline the process.

The bottom line? You need someone on your team who will own the data and look after the way you source, capture, integrate, clean, validate, enrich, master, deploy, and govern it.

If you've never wrangled data, all of this will sound daunting. If you have, you know that it's really common sense combined with clever tools. If you're leading a Data Lake initiative, you don't have to be the one to manage all the data. But you will need to find someone who can do this – and give them all the things they need to do the job properly.

“A basic truth of management – if not of life – is that nearly everything looks like a failure in the middle.”
Rosabeth Moss Kanter

Our 60-day sprint

So our foundations were solid: robust implementations of analytics, marketing automation and CRM. And our data was cleaned, mastered and ready to play with. Now it was time to make it all happen. In our 60-day program, we promised ourselves, we would do four important things:

- **Connect our remaining silos**, so the data from each marketing app we use would integrate with the others (a number of these we had in place before we started the sprint).

- **Put the neatly linked data from all the apps into our Data Lake**, where we could slice it up, put it to work, and run our weird and wonderful queries.

- **Stick a visualization tool on the front**, so we could create new dashboards, reports and drill-downs.

- **Introduce our new capability to our colleagues in Sales**, so they could put this new insight to work and think up new, revenue-driving use cases.

Linking core marketing applications via APIs

The first two weeks of our sprint were focused on creating the data linkages across our core marketing applications using the Application Program Interfaces (APIs) that each app provides.

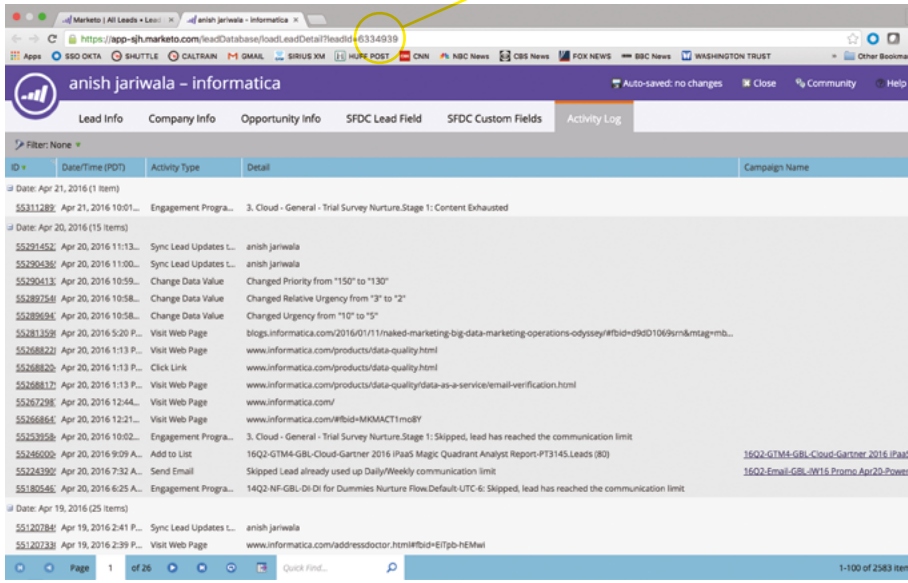
APIs are the critical connection points of the digital marketing era. They allow any application to exchange information with any other. A tech vendor's success often depends on the robustness of its API. APIs matter.

For our 60-day sprint, we concentrated on using APIs to ensure the following five integrations work:

1. The Adobe Analytics and Marketo link
The most important data link for us is the link between Adobe Analytics and Marketo.

Every new visitor to our website gets two IDs: the Marketo ID

↓
fig. 10
Marketo: Lead Database → Activity Log.



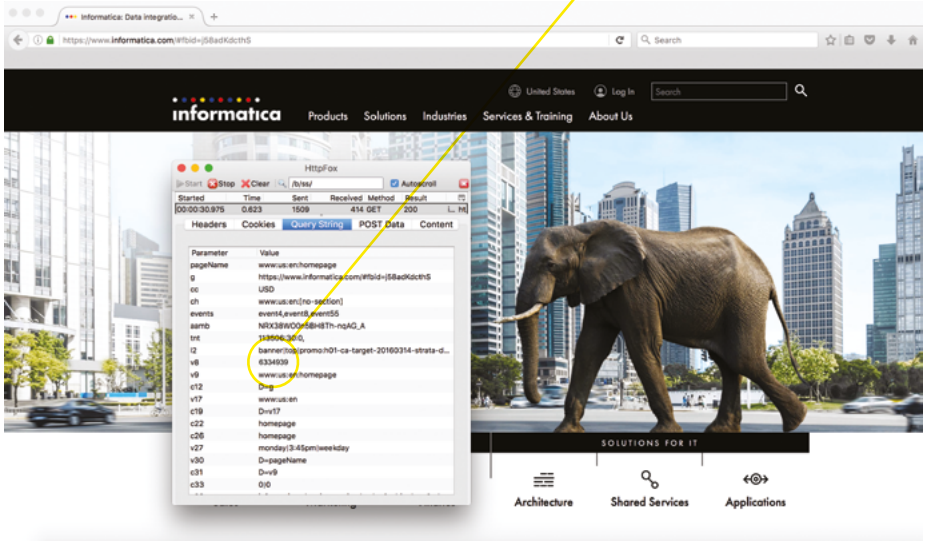
and an Adobe Analytics ID. All of each visitor's activity is tracked in both, initially as an anonymous user. Once the person fills in a form, all of his or her past activity immediately accrues to the Marketo ID, making them a known visitor.

Fig. 10 is the Activity Log for Anish himself in the Marketo system, showing all of his activity, with a timestamp.

Clearly, this data carries important indicators about Anish's intent, product interest, etc. We needed to expose this Marketo ID in Adobe Analytics – making it the key to connecting these two data sources in our Data Lake. So after exposing Marketo's SOAP and REST APIs to our website using Adobe Tag Manager (more on this below), we were able to capture the Marketo ID in Adobe Analytics as a dimension (this is an eVar in Adobe's lingo).

Fig. 11 is the screenshot of the Adobe Analytics tag that fires when someone visits our website. As you can see MKTO ID 6334939 is captured under the dimension (eVar8 – v8).

↓
fig. 11
How Adobe Analytics eVars and Success Events are captured.



That innocuous-looking entry is actually a big deal. It means that we can connect all of Anish's activity across the website with all of his other activity as he engages with our campaigns (his email opens and clicks; his webinar registrations; his visit to our booth at Informatica World… you get the idea).

Signal from noise

But what if Anish visited our website from different devices? What if he used his laptop at work but an iPad at home?

The bad news: Adobe Analytics and Marketo both generate a unique ID for every new browser per device. So, until Anish fills out a form on another browser (device), he looks like two or more different people to us.

The good news: Marketo has a de-duping feature called 'Merge Leads'. So when Anish uses the same email address to submit a form on different devices, Marketo collapses the info into the original Marketo ID – very, very handy.

In the Adobe Analytics screen shot (fig. 12), you can see that Anish's Marketo ID – 6334939 has five Web visitors associated with it – he used five different browser/device combinations to access our website but submitted forms from each using the same email. So we've got all five devices connected to Anish now (fig. 12).

If the same visitor uses multiple email addresses to submit a form, Marketo will always assign a new ID every time, concluding that they're different people. Right now, we can't do anything about this noise – but we estimate that it only applies to less than 10 percent of our data.

If this were a financial system, we'd need to reduce this noise to zero. For a marketing system, we're happy to have reduced the noise from the previous 40-50 percent of our data to just 10-15

↓
fig. 12
Adobe Analytics: Custom Conversion → Marketo ID (Evar8) and traffic metrics.

percent. We can still make good decisions about who we should call or target or include in a campaign with this level of accuracy.

2. Adobe Analytics and Demandbase integration
Demandbase uses reverse IP-lookup to provide firmographics like industry, company name, and location whenever someone visits our site (as long as they visit from their business office or when they're connected to their company via VPN).

Demandbase has various APIs that integrate with Adobe Analytics – it only took a few hours. We just had to decide which dimensions (eVars) to use to populate Adobe Analytics with our Demandbase data fields and we were all set. Again, just as with Adobe Analytics, all Demandbase information is anonymous until we connect it to a known lead in our Data Lake.

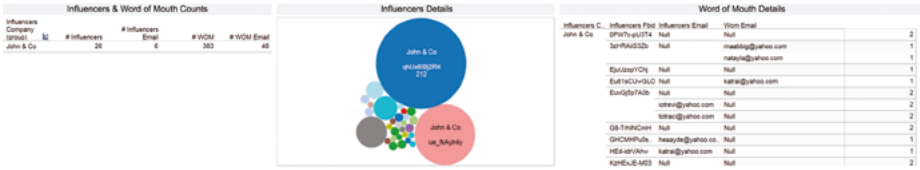
3. Adobe Analytics and Rio SEO integration
We use Rio SEO to capture social shares, influencers, and word-of-mouth activity (people coming to informatica.com by clicking a link in a social share, content shared via email or chat, etc.).

About 1 percent of our traffic comes from people we call influencers – people who share our content with their colleagues and followers. This influential one percent brings in a whopping 5-7 percent of our site traffic. And we believe this to be some of the most important site traffic we're getting.

When someone uses one of our share buttons on the site – or when they copy/paste a URL into an email – we can identify the original sharer when the recipient clicks on the link.

Rio SEO has its own unique code called FBID (it's device/browser

↓
fig. 13
Tableau dashboard (detail).



specific so you can call it a Rio SEO cookie). This ID is appended on every URL when you visit informatica.com. Look up at the address bar on our site and you'll see yours.

For Anish, his FBID is: j58adKdcthS. We capture this as eVar 60 in Adobe Analytics and you can see it in his profile (fig. 14).

When Anish sends a link to others, they get their own unique FBID when they click on the link. They become the WOM – Word of Mouth – and he is the influencer.

We receive Rio SEO data via FTP on a weekly basis, showing the influencer/WOM relationships. We match this information in our Marketing Data Lake and surface the relationships in our Tableau dashboard.

All these integrations described so far are for anonymous visitors – Adobe Analytics, Demandbase, and Rio SEO. These relationships and analytics are great to have and certainly help inform our marketing but if I were to approach our sales team with this information, they'd want company and contact details. That's why making connections in the Data Lake is so important.

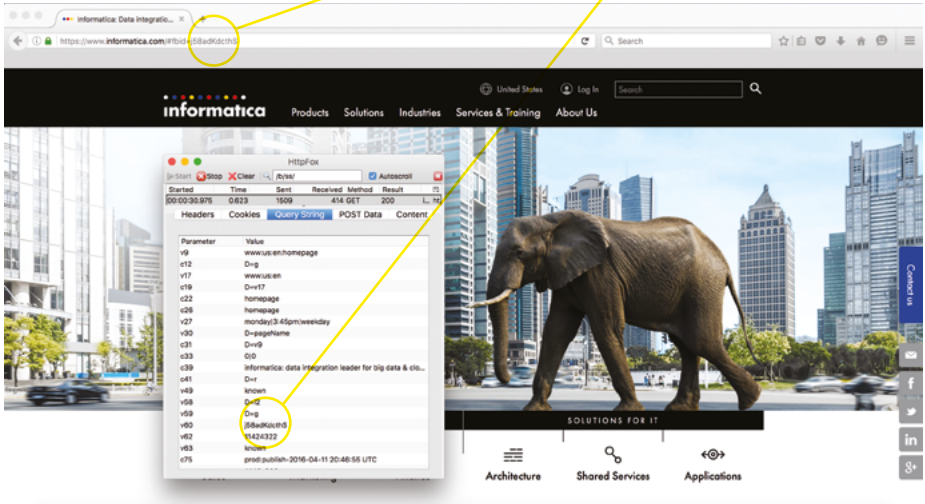


fig. 14
How Adobe Analytics eVars and Success Events are captures, image showing eVar (v60) capturing FBID.

4. Marketo and Lattice Engines integration
Lattice is the predictive scoring tool that we talked about earlier. Based on historical patterns it tells us how likely a prospect is to buy from us – a hugely important guide for our sales and marketing efforts.

Marketo and Lattice Engines integrate really well. Every time a lead is created in Marketo or a key activity is updated at a lead level, the lead record is scored or re-scored based on one or more of our Lattice models (we built different ones for different product lines and geographies) – the scores are directly written to Marketo via API.

The nine activity types we use to trigger Lattice scoring are:

1. New Lead
2. Click Link
3. Visit Webpage
4. Interesting Moment
5. Open Email
6. Email Bounced Soft
7. Fill Out Form
8. Unsubscribe Email
9. Click Email

There are about a thousand different variables and attributes (available via Lattice) that can be included in the predictive models in addition to the triggers we provide – things like technology profile, website profile, firmographics, website keywords, growth trends, etc. Of course not all them will turn out to be positive or negative predictors for conversion to revenue, but the ones that are will become part of the predictive model.

The Lattice score comes back into Marketo then on into the Data Lake. The rating that's stamped at the point of lead to opportunity conversion is ultimately pulled from the Salesforce opportunity object and loaded in. That means the Lattice models are always learning and improving over time.

5. Marketo and Salesforce integration

Salesforce holds all the data about what happens to leads once they're passed to Sales. This, obviously, is critical for modeling our world and achieving the end goal: attributing revenue to marketing touches.

Fortunately, Marketo and Salesforce have a good basic bi-directional API sync. Every 5-10 minutes, fields that are updated in either app get synced in both directions, Marketo to Salesforce, and Salesforce to Marketo. Data, especially custom fields, that are not covered via the basic integration or require some additional massaging are synced using our own Informatica Data Integration tools – the advantages are that there's no coding required, debugging problems is simple and any issues get reported. When a new record is created in Marketo (through a form-fill, a list upload, the web API or manually), we had two options:

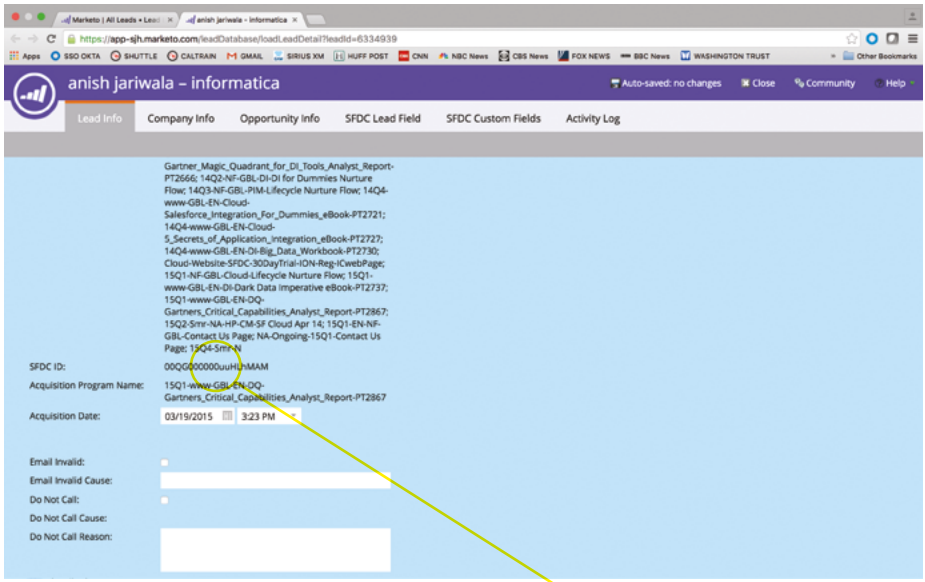
1. Update a record in Salesforce in the next sync cycle.

2. Update a record in Salesforce when the lead is considered ready to be contacted by Sales.

At Informatica, we decided to follow option #1 – syncing a lead to Salesforce right away to give visibility to Sales. Complete transparency. A lead will move to Salesforce from Marketo only if the minimum attributes are available. For us, this includes first name, last name, company and email address.

In the reverse direction, a Lead or Contact record will sync from Salesforce back to Marketo as long the record has an email address. Then the Salesforce ID is passed over to Marketo when the lead syncs.

Below (fig. 15), you can see the SFDC ID for Anish in the lead record in Marketo. This SFDC ID is the glue (the connecting key or join) between Marketo and Salesforce.



↓
fig. 15:
Marketo: Lead Database → Activity Log. Connecting Marketo to Salesforce using SFDC ID as a key.

What all these data connections mean

These five bi-directional data handshakes are actually much more than that. Because all of this data lives in our Marketing Data Lake, the connections we created between the data sets in Adobe Analytics, Marketo, SFDC, etc. mean that ALL THE DATA IS NOW JOINED UP VIA A SINGLE LEAD OR ACCOUNT ID.

No more silos.

Slapping a pretty face on all this

Okay, so all this data is in the Data Lake. But to gain insight, we need to expose it to the people who can put it to work: the field marketers, SDRs, biz dev teams and marketers who are making decisions about who to target, what to say and what tactics to use.

That means a dashboard – or a series of dashboards to be more accurate. A dashboard is the thin layer that sits on top of our Data Lake, a window, showing us various views about what's in it.

It's funny that a lot of people in the company are now calling what we do 'The Tabelau Dashboard' as if the entire project was just a matter of sticking a dashboard on top of some clumps of data – purely a visualization exercise. Unfortunately, this is how a lot of teams buy their data visualization tools too: as if the last mile of data quality, integration and analytics is actually the whole journey!

Tag management

We already mentioned Adobe Tag Manager but we can't talk about all this integration work without giving it a hat tip. ATM, like the Google Tag Manager, is a tool that helps you manage all the JavaScript and tags that you deploy on your website in one place.

Without it, we'd have to manually place the various code snippets on every relevant page for every marketing app we use (from web analytics to marketing automation to social shares… all of them). Then, whenever we wanted to change the code, we'd have to go back to every instance of that snippet and update it manually. That's a huge hassle that would prohibit a lot of what we're doing.

With ATM, you inject and update all scripts and tags in one convenient place, and do it once. Which is a lifesaver. And it works for more than the tags we discussed here e.g. advertising, retargeting tags, etc.

The Aha! moment

About 30 days into our 60-day sprint, we had an "Aha!" moment – a moment when we knew what we were doing would really add value to the whole revenue generation machine.

For us, it was when we saw a real Marketo ID for a specific web visitor combined with our Adobe Analytics data. We realized that we now had the email and phone number of that visitor associated with all his web activity. Up until that moment, all of this was just theory. From that moment on, it was our new reality: a connected sales and marketing operation that linked all of our data to a profile; an end-to-end view of the whole customer journey.

We'd like to say we heard the sound of angels singing or that we popped open a bottle of champagne to celebrate. The truth is, we said, "Nice one." And went home. There was still a lot of work to do.

Five lessons

So that's a quick tour of our 60-day integration sprint. On Day One, we had a collection of marketing applications and platforms but no way to connect them. On Day 60, we had a joined-up revenue engine with full visibility into customer journeys and marketing programs. Here are five lessons we learned on the way:

1. Start your journey with a business outcome in mind.
A lot of teams start journeys like this with a list of questions they need to answer or reports they'd like to see. That's okay as a rough guide, but it could paint you into a corner.

Instead, start with the business outcomes you want to drive. The questions you want to answer will almost surely change over the course of the project – and might even get in the way of achieving the real business value. So start with goals about real outcomes, like lead-to-opportunity conversion rate increase or improved SDR outbound call close rates. And keep all eyes on the prize.

2. Start with the revenue side.
We're all here to sell stuff. So make sure your Data Lake program prioritizes delivering relevant insight to the sales team. Budget optimization in the marketing department is a great thing. But what really motivates sales people is when they see real new

opportunities and can drill down for the insight that creates sales conversations. We're glad we started there.

3. Make friends with IT.
There's no way our marketing team could do all this integration without the active involvement of our IT colleagues.

But IT is charged with delivering the most value to the business for the least investment. Until we sat down with them, their natural reaction to our plan was, "We have most of this data in our data warehouse. We don't need yet another data store." They had a point: why duplicate data?

Once we'd talked them through our needs and strategy (and showed them how inexpensive it is to hold this data in Hadoop) the IT team didn't just accept our program, they were enthusiastic about it.

Soon, the challenge wasn't to get IT people to help – it was how to make everyone else in IT feel a part of the project. As a group, we made sure that everyone learned about the project, even if they weren't driving it. So everyone in warehousing is also getting Data Lake and Hadoop skills.

And just as you can't do this without IT – IT can't do it without you. You need to be actively engaged in translating your business logic over to the IT people so that they can integrate things in the right way. This is a two-way commitment.

4. Make sure your web analytics and marketing automation work well independently before you bring the data together.
You need to know that each system is doing exactly what you need it to do before trying to sync them up. To get an idea of what we mean by that, review the Checklist chapter.

5. Market your successes.
It's important to let everyone know what you're doing and how it's working. We gave frequent updates to key stakeholders in Sales and Marketing but also to the CEO and even the company's board. The high visibility and clear communication about goals and

progress turned the project into a runaway success that earned all the attention and help we needed to maintain the momentum.

“Traditional demand generation is fishing with nets. Account Based Marketing is fishing with spears.”
Jon Miller, Engagio

Account Based Marketing and the Data Lake

B2B purchasing is a team sport – especially when it comes to the biggest deals. Rarely does a single individual make the entire purchase decision for an enterprise solution. Instead, a multi-disciplinary buying team, with anywhere from 5-12 people, collaborate on the decision, each stakeholder making sure the solution will work for their department as well as the entire organization.

That's why sales teams have always been organized around accounts. And the bigger the deal, the more of a tailored, account-specific approach is required.

But most marketing technology today takes a lead-centric view of the world, tracking individual contacts with no view into their membership of a company or team. Account Based Marketing is a response to that.

Account Based Marketing (ABM) is the systematic orchestration of all marketing and sales efforts, focused around a limited list of target accounts.

It's a powerful concept that brings Sales and Marketing closer together than ever before. But it demands something that isolated marketing apps do not provide: an account-based view of the world, with visibility into how specific accounts are engaging with your website, content and marketing programs.

The Marketing Data Lake is a powerful tool for solving the visibility challenges of Account Based Marketing. In fact, the very first use case that we decided to tackle for our Data Lake program was creating an account-based dashboard for our sales and business development teams (see the Dashboard chapter).

The goal: total account visibility

The end game is to be able to show our sales teams exactly what's happening inside the accounts they care most about:

> Who in the account is engaging with us.
> Where they're engaging.
> What products they're interested in.
> When that engagement spikes and when it dissipates.
> Who is on the buying team and how they relate to each other.

If we can even start to answer these questions, the productivity and effectiveness of the sales team should improve dramatically (and that's exactly what's happening). Why? Three reasons:

1. Sales Development Reps can focus their efforts on accounts that show the most promise.

2. When they do get in touch, the SDRs can have more relevant conversations that the prospect is more likely to welcome and value.

3. Otherwise invisible opportunities within a given account get exposed early, so sales people can cross-sell.

Any one of these uplifts is a business case in itself. Taken together, they're irresistible.

How the Data Lake helps

The Data Lake supports ABM by combining data from different sources to generate an account-based view of cross-channel marketing interactions. A single marketing application, like Marketo, can give you a lead based view of the data inside it. But for a multichannel view, you need a way to connect systems while preserving the account ID.

With our Data Lake, we can associate prospects with their accounts and with the other prospects inside those accounts. We do this even for anonymous visitors, using the Demandbase reverse IP-lookup discussed earlier.

If a prospect visits our website from an IP address that Demandbase recognizes, the company (and its associated firmographics) is attached to his or her profile in the Data Lake. Even when they're still unknown to us.

The match rates for this runs at around 30-40 percent, so we know we're not yet getting the most complete picture. But the Data Lake makes it easier to supplement our knowledge over time, using more data sources to fill out our view. (Of course, when the prospect fills out a form, we can use their email address and/or the Company field to link them to their account.)

From that moment on, all of their activity is tagged as being associated with that account. So the account page on our dashboard can aggregate the activities of everyone in the company, in a single view.

We also have some added tactics for spotting who in each team is sharing content with whom – to spot the influencers and the influenced.

A little insight is better than none

We've only started out on our journey to create ever-richer and more up-to-date account profiles using our Data Lake. But the benefits we're already seeing even with our partial view are remarkable.

If that's in any doubt, look at the interview with two of our sales leaders. They and their teams keep the account based dashboard open at all times as they go about their work, looking for new opportunities, tuning existing ones and turning cold calls into informed, relevant conversations.

Now let's take a look into the dashboard itself, with special attention on the ABM page.

“Sustained success is largely a matter of focusing regularly on the right things and making a lot of uncelebrated little improvements every day.”
Theodore Levitt

The Data Lake dashboard

As you've seen, a lot of the Marketing Data Lake approach comes down to data plumbing. It's about integrating your different data sources into a single place; maintaining the quality and integrity of that data; and creating a golden record for each prospect as they progress on their journey to revenue.

But all of that data plumbing is worthless if it doesn't make an impact on the people who actually sell things: the business development managers, field marketers and sales people charged with generating the revenue that keeps us all in our jobs.

That's why our first use case for our Marketing Data Lake was a dashboard that gave our SDRs and sales people insight into the people they should be calling, in the accounts they cared about – and insight into what to say to them once they called.

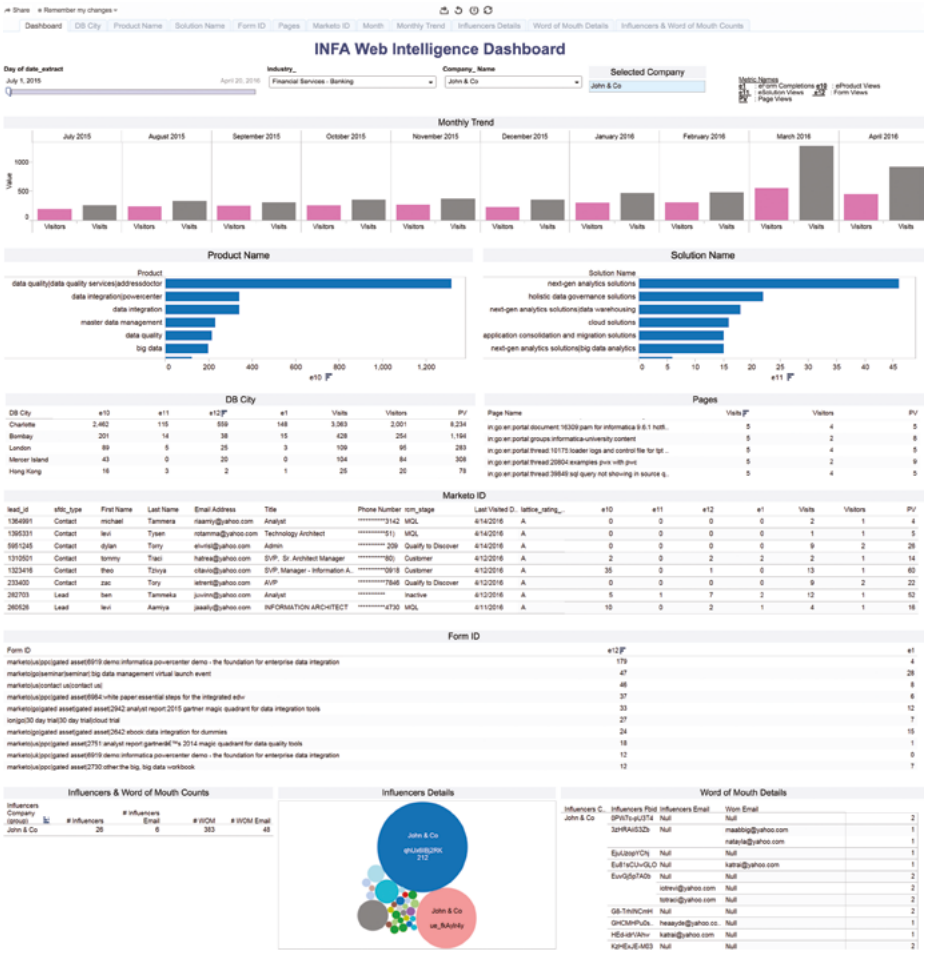
In the spirit of opening our kimono and sharing everything we do, we're going to show you something you'd normally need a password and two-factor authentication to access (all the data in the screen grabs is anonymized using our data masking tool to protect… well, to protect us).

Our dashboard

All dashboards pretty much look alike: a cluster of bar charts, tables and timelines. Here's ours, created and continuously improved by Anish (fig 16).

The magic, of course, is in the actual data contained in all these nifty widgets. So let's go on a tour.

↓
fig. 16
The home dashboard: An account view of all lead activity for a company visualized via our Data Lake.



The tabs

Our dashboard is actually several dashboards, each with its own tab. We've opened the main tab, the home dashboard, but you can see we've got different ways to slice our data, each with its own tab and sub-dashboard.

For instance, 'DB City' is a view of the activity within any given city (people from that city visiting our site or interacting with our campaigns). DB stands for Demandbase, our reverse IP-lookup provider.

The other tabs are pretty self-explanatory: a view by Product, Solution, Form Fills, Web Pages, Marketo ID, etc.

Now for the main sections of the dashboard:

1. An ABM view

↓
fig. 17

The top slice of the main dashboard shows what company we're looking at. That may not seem like a big deal, but it actually changes everything, turning our lead-centric marketing systems (CRM, marketing automation and web analytics) into an account-based view.

Account Based Marketing is a hot topic and major growth area within B2B marketing and for good reason. It's based on the simple premise that the biggest B2B purchases are always made by buying teams, not by individuals acting alone.

Getting a clear view of an entire account – and how the people in that account are interacting with you – is an incredibly powerful tool for closing the biggest, most important deals. It makes Account Based Marketing possible. So the rest of this chapter will be about this view, the account view.

By the way, the upper right-hand corner of the dashboard is the key to the measures we use throughout the dash. So e1 stands for Form Completes and e12 for Form Views, etc.

2. Visits and Visitors

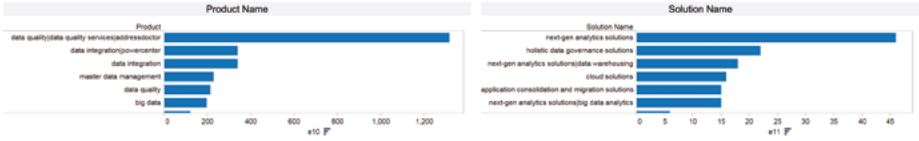
↓
fig. 18

This slice shows the monthly visits and visitors coming from the selected account, over time. The data comes straight from Adobe Analytics, having already gone through the Demandbase filter to establish the company and industry.

The visit count is always higher than the visitor count because a single visitor can come back many times.

Many of these visitors are still anonymous but some are known, because they filled out a form somewhere. Importantly, we can still associate the anonymous visitor with his or her company, so they're still included in this account-based view.

3. Product and solution interest



↓
fig. 19

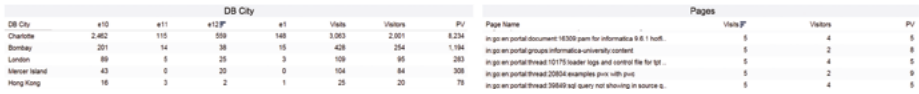
This horizontal bar chart shows the products and solutions that the people in this account are showing the most interest in. It comes from the web analytics: what pages people looked at, what content they consumed and what forms they completed.

Again, this is hugely important to our sales teams, helping them

spot new opportunities within an account that they may not have known about. Our sales people tell us how powerful this has been for expanding opportunities and increasing cross-sell (see the Q&A with our sales leaders).

→ **Drill Down: by product**
With a click, we can also look at our world from a Product Interest or Solution Interest perspective, showing all companies and individuals with an interest in our MDM product or our Data Quality solutions. When we do this, the entire dashboard and all data in it is filtered by the product selected.

4. Location



DB City

DB City	e10	e11	e12	e1	Visits	Visitors	PV
Charlotte	2,462	115	559	148	3,063	2,001	8,234
Bombay	201	14	38	15	428	264	1,194
London	89	5	25	3	109	95	283
Mercer Island	43	0	20	0	104	84	308
Hong Kong	16	3	2	1	25	20	78

Pages

Page Name	Visits	Visitors	PV
in:go:en:portal:document:16308:pam for informatica 9.6.1 hotfi..	5	4	5
in:go:en:portal:groups:informatica-university:content	5	2	8
in:go:en:portal:thread:10175:loader logs and control file for tpt ..	5	4	5
in:go:en:portal:thread:20804:examples pxix with pvxi	5	2	9
in:go:en:portal:thread:39849:sql query not showing in source q..	5	4	5

↓
fig. 20

Most big companies have many locations. Being able to spot the location of the visitor interacting with us helps us spot likely buying teams and centers of interest within a bigger company.

The list on the right shows the web pages visited from each of the locations identified. The page names are human-friendly (instead of gnarly URLs), so we can easily see the top interests at each location.

We've often heard the sales people say things like, "We knew about our Data-as-a-Service deal with the folks in Lexington but we had no idea they were also looking at our Cloud Integration tools."

→ **Drill Down: by territory**
We can also drill down into any territory, looking at all accounts and individuals within that area and their interests.

Our sales team is organized by territory, so that's a critical view for them.

5. Activity of known prospects

Marketo ID

lead_id	sfdc_type	First Name	Last Name	Email Address	Title	Phone Number	rom_stage	Last Visited D..	lattice_rating_..	e10	e11	e12	e1	Visits	Visitors	PV
1364991	Contact	michael	Tammera	ntammy@yahoo.com	Analyst	**********3142	MQL	4/14/2016	A	0	0	0	0	2	1	4
1365331	Contact	levi	Tysen	rotamma@yahoo.com	Technology Architect	**********151	MQL	4/14/2016	A	0	0	0	0	1	1	5
5951245	Contact	dylan	Torry	ehrisk@yahoo.com	Admin	**********209	Qualify to Discover	4/14/2016	A	0	0	0	0	9	2	26
1310601	Contact	tommy	Traci	hatrea@yahoo.com	SVP, Sr. Architect Manager	**********80)	Customer	4/12/2016	A	2	0	2	2	2	1	14
1323416	Contact	theo	Tziya	olavio@yahoo.com	SVP, Manager - Information A..	**********0918	Customer	4/12/2016	A	35	0	1	0	13	1	60
233400	Contact	zac	Tory	ierrem@yahoo.com	AVP	**********7846	Qualify to Discover	4/12/2016	A	0	0	0	0	9	2	22
262703	Lead	ben	Tammeka	juvinn@yahoo.com	Analyst	**********	Inactive	4/12/2016	A	5	1	7	2	12	1	52
260526	Lead	levi	Aamiya	jaaally@yahoo.com	INFORMATION ARCHITECT	**********4730	MQL	4/11/2016	A	10	0	2	1	4	1	16

↓
fig. 21

Seeing the specific contacts within each account and their activity is a hugely powerful part of the dashboard. This is one of the key places where the integration between Marketo, Salesforce and Adobe Analytics comes in.

The table shows:

- Name and Marketo ID
- Whether they're a lead or contact in Salesforce
- The date of their last web visit
- Their Revenue Cycle stage
- Their Lattice Rating (predicting their likelihood of purchase)

You can imagine how much this helps guide our sales people to the most interesting prospects – and informs their conversations.

6. Content consumption

Form ID

Form ID	e12	e1
marketo(ua)ppc(gated asset)6919 demo:informatica powercenter demo - the foundation for enterprise data integration	179	4
marketo(go)seminar(seminar) big data management virtual launch event	47	26
marketo(ua)contact us(contact us)	46	8
marketo(ua)ppc(gated asset)6864=white paper:essential steps for the integrated edv	37	6
marketo(go)gated asset(gated asset)2642 analyst report:2015 gartner magic quadrant for data integration tools	33	12
ion(go)30 day trial(30 day trial)cloud trial	27	7
marketo(go)gated asset(gated asset)2642 ebook data integration for dummies	24	15
marketo(ua)ppc(gated asset)2751 analyst report:gartner's 2014 magic quadrant for data quality tools	18	1
marketo(ua)ppc(gated asset)6919 demo:informatica powercenter demo - the foundation for enterprise data integration	12	0
marketo(ua)ppc(gated asset)2730 other:the big, big data workbook	12	7

↓
fig. 22

This table shows which of the gated content assets on our website were viewed by the people in this account.

The first column shows the content asset, the second shows visitors who viewed the form (e12) and the third column the actual form fills (e1). We expect a certain amount of bouncing away from any page with a form on it – but with this view, we can still capture their interest in the content even if they chose not to complete the form.

Again, content interest and consumption are great insights for sales conversations.

7. Influencers

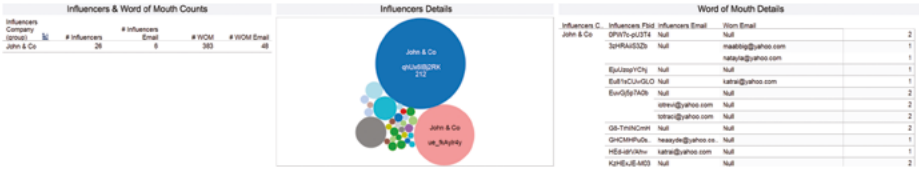


fig. 23

This section of our marketing dashboard identifies the most influential people within the account – the people who are sharing links and content with their colleagues.

We use Rio SEO to attach a unique code to the URL of every web page viewed. When that URL is cut and pasted into an email or shared in social media, we can associate the subsequent visit by the receiver of the email or the person clicking a link in social media. That means we can see where our word-of-mouth power is coming from.

Imagine this dashboard in your own company

So that's a quick tour of the first use case for our Data Lake, our Account View dashboard.

When our sales people first saw this, they used the term 'game-changer' a lot. And it really is: for the first time, Marketing is helping

Sales spot real opportunities early and guide the conversations about those opportunities. That means they spend significantly more of their time on deals that are likely to close and significantly less time on cold calling.

Here's where we turn into a cheerleader. Because we don't want you to just look at our dashboard and say, "Wow. That's cool." We want you to make one of these for your own company.

Everything represented here is within reach of every B2B company today. The blueprint is here in this book. All you need is the will to make it happen. We're completely convinced that this is the way B2B marketing has to go – and that if you pursue it, your company will sell more and your career will take off.


Q
&
A

Q&A: Our sales leaders

So far, this book has been told from the perspective of the marketing team. In this chapter, it's time to get the perspective of the people we do it for: the people who turn the opportunities into revenue.

One of the most gratifying parts of the whole Data Lake journey has been the dramatic change in the relationship between Marketing and Sales. Today, it feels like we're part of one, completely aligned team. With mutual trust, accountability and total transparency into each other's processes, activities and data.

What does that look like from the Sales side? Here's a Q&A with two of our most senior revenue generators:

Scott Padlick, Senior Director of North America Field Marketing
Scott is responsible for generating pipeline for every sales rep in the US and Canada.

Aaron Klein, Director of Business Development
Aaron runs the enterprise & corporate business development team.

These are the guys who run the teams that turn potential into money. They're the people we created our big data marketing platform for. Let's find out what they think of it.

Q What was selling like before the big data marketing program?

Scott If you go back a year, our prospecting was very cold. It wasn't personal. We'd work with the sales team to identify some top accounts, load the names up for Aaron's business development team and they would make cold calls. Essentially, we were trying to find a needle in a haystack.

Aaron The way we did it a year ago was very inward out – our conversations were about features and products. It was a lot more 'spray and pray'.

Q So has the new program changed that?

Scott When we got the Data Lake dashboard, we could instantly adopt a more intelligent and personalized approach.

Now we take those target accounts and look them up in the dashboard. We can see their web visits; we see exactly who the individuals are and what pages they're looking at.

So the business development team is able to be much more selective about the people they reach out to and much more personal in their approach because they have context about what those people are interested in. That's a huge difference.

Aaron Our conversations are now more about impact on the business and building a true relationship as a trusted adviser – not just a rep trying to push a product.

It's a lot less spray and pray and more of a focused approach to our account sets – and that's made a huge impact for us.

Q Does the platform also help spot new opportunities as well as accelerating existing ones?

Aaron Absolutely. One way is that it shows us additional products of interest in a prospect account – something we didn't know they had an interest in. We can see a team that we may be talking to about, say, MDM, and spot that they're also hitting a lot of web pages and content about Data Quality. So we can open up that opportunity.

Scott It's insight that can spot new opportunities, grow them in new directions and increase the average sales price.

Aaron We just recently we saw a huge spike in interest on an important account – a 300 percent increase in web traffic and activity in 30 days. That says something is happening within that account that needs our attention immediately. That's an opportunity we would never have seen without this new approach. This tool let us spot the new spike and accelerate that deal.

Q How does the new insight help in sales management?

Scott The sales managers working with Aaron and his team also use it as a coaching tool.

So if the sales managers hear from a rep that he or she has pursuits around a given product in a handful of accounts, we can go into the dashboard and see if those people are hitting those product pages on the website. If they are, that's great. We're breaking through.

And if we're not seeing web activity, the manager goes back and works with the rep to make sure we're getting to the right people – to further qualify those opportunities.

Q How do you use the predictive scoring?

Aaron This is big. The Lattice Engines tool allows us to score the quality of the contact or the account according to their propensity to buy.

It goes way beyond looking at their last visit to the website. The score is based on hundreds of attributes and behaviors and it lets us determine the quality of an opportunity.

The idea is to focus our team's efforts by asking, "Who do we go after?" instead of simply going after everyone in the system as we did in the past.

It's right there in the dashboard, so you can see, "Wow, someone has hit our MDM page twelve times and, look, they're on Lattice A."

Q What about the 'influencers' piece? How does that help?

Scott Not only are we able to see who's visiting the website and what they're looking at, we can also see who in that buying team are the most engaged – the people who are sharing information with their colleagues and who seem to be the biggest advocates for Informatica.

So now the business development team knows specifically who they should be reaching out to, with a very targeted personal approach with lots of intelligence and context.

Q Marketing has a history of bringing shiny new things to the table. Was there a lot of skepticism about this new big data approach?

Aaron To be honest, when I saw this I actually started salivating because this is something that I'd never seen before.

It gives us true insight in real time about what's happening – and that's dramatically different from the kind of pie-in-the-sky stuff you're talking about.

This is real. When you get your hands on it and you have a business conversation with a contact, you know this is live and active.

Scott The feedback from the field sales reps and the sales managers has been the same.

Aaron Yes. Every company has their skeptical reps and even the skeptical reps or managers have said, "Okay, this is pretty cool. I can see how that information would be valuable."

Scott Actually, I'll even quote a Regional Manager and a VP who said, "Holy <beep>. This is amazing." One guy manages a team of 8-15 field reps and the VP runs an entire region. So, for them to actually make that comment to me was very, very, encouraging.

Q Okay, give me the bottom line. Summarize this new way of working.

Aaron It really is a game changer. Fewer than one percent of the companies in the world are using the framework that we deploy. More importantly, it's going to be a key factor of our growth from 1 to 3 billion over the next 3 to 5 years.

Scott Put it this way: I'm in the tool every day. I bounce around from region to region and look at opportunities that are currently in flight, that we've forecasted to close, and then at our white space and those accounts that we're targeting.

I use it to determine where there's a significant influence or an up-tick of interest and making sure the teams and managers are fully aware of that. So I'm in the tool daily. Knowledge is key and my goal is to put as much in front of our reps as possible so they can target the right people in the right way.

A new relationship

We've been in B2B marketing for many, many years and this is the first time we'd ever heard sales people talk this way about a marketing-led initiative.

For us, it's not just exciting because of the results (that would be more than enough). It's also exciting because it shows a new way forward for the most important relationship in any business: the relationship between sales and marketing.

With the Data Lake approach, we can align the two teams more closely than ever before – and do it around a clean, clear data set that shines a bright light on real opportunities. It's the difference between an inefficient, broadcast-style model and a focused, targeted effort built on relevant conversations with real prospects.

And that's what gets us out of bed in the morning.

“Does your content lead readers on a journey, or does it merely stuff them as leads into a pipeline?”
Ann Handley, Marketing Profs

The role of content

If our new marketing operations stack and processes make up a revenue engine, then content is the fuel.

Before our prospects ever come to us to learn about our products, they come to educate themselves about the issues that matter most in their work. Our content offers – from blog posts and videos to technical workbooks, surveys and analyst reports – are what our prospects come to us for.

The idea is simple: package up your expertise to help your prospects do their jobs better and good things happen:

- You engage with them earlier in the purchase journey.
- You educate them, helping them make better choices.
- You position yourself as experts who are motivated by the success of your customers.

– You pick up important signals about their intent, product interest and readiness to buy.

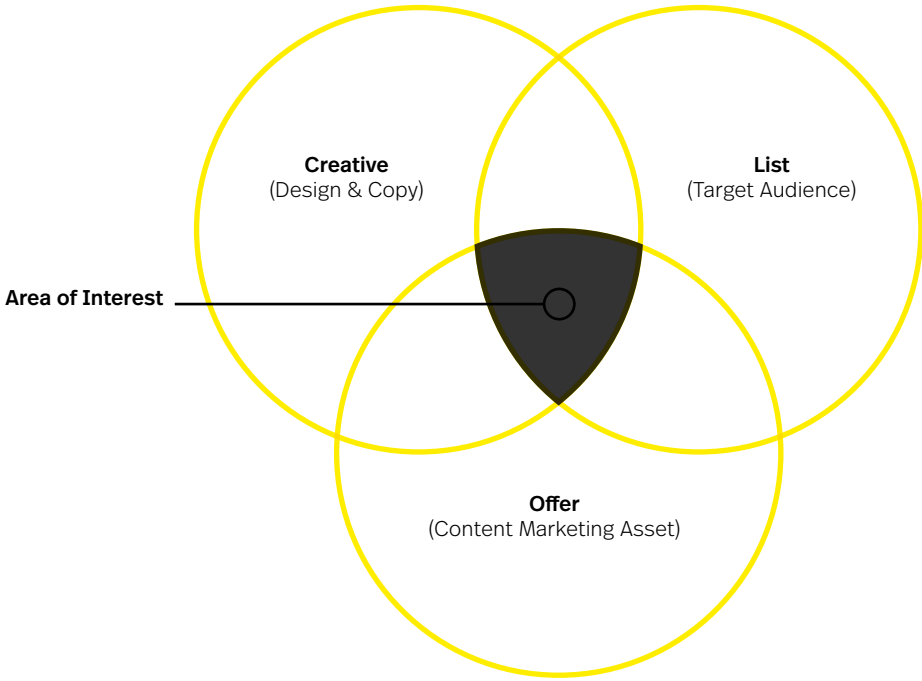
– You can track them as they progress along the funnel, until sales-ready.

– You bring value to conversations across social media and industry forums.

– You earn traffic when people are searching for the kinds of solutions you sell.

The new B2B buying process is self-guided, search-initiated and social media-informed. The days of pushing product messages into everyone's faces are long gone.

↓
fig. 24



But as much as things have changed in the new marketing, some fundamentals remain unchanged. Success still depends on getting three things right: the list, the offer and the creative. In our new model, the list is tighter and better segmented; the creative is continuously improved through testing; and the offer is a growing library of super-relevant content. That's a killer combination.

You don't need us to convince you of the merits of content marketing here. But there are specific considerations to keep in mind when working with content in a Data Lake environment.

To gate or not to gate?

When people first plug in their marketing automation software, there's an enormous hunger for names and email addresses to start pouring into the top of the funnel – something real to track as it moves through the machine.

As a result, many marketing teams tend to gate every single piece of content, putting it behind a big, fat web form. But when you look at the data, you tend to see enormous bounce rates from pages with forms on them. So for every person you entice into filling out a form, there are dozens or even hundreds that say, "No thanks," and click away.

That hurts. It wastes all the money you invested to get them to that page. And it misses a big opportunity to engage with, educate and earn the trust of a potential prospect. Gating also hurts you in your SEO rankings as too much content is hidden away from the search engine spiders.

The more we looked at the data, the fewer forms we deployed. When we migrated our site to Adobe Experience Manager, we had over 1,200 forms. Today, we have 300 and we plan to have even fewer. No top-of-funnel content is gated; most mid-funnel content is; and for bottom-of-funnel content, it depends on the situation. This approach is better for SEO, it delivers a better user experience

and, in the end, we actually get more form fills from higher-quality prospects. How can we tell lead quality? From our predictive lead scores; the question is simply “How many A-leads is a channel, program or content-offer driving?”

Test everything

We haven’t discussed testing very much in this book because it’s so fundamental to best practice digital marketing that we assume you’re doing it already.

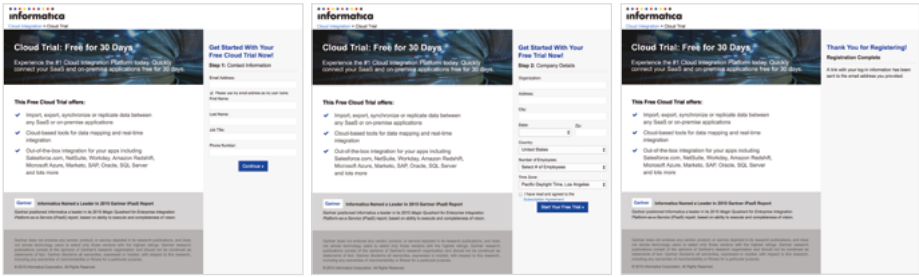
When it comes to content, testing is even more important. So we run a continuous program of tests (We run mostly A/B tests. Depending on traffic levels for the page to be tested, we may not have the patience for multivariate testing – it simply takes too long to get to statistically significant conclusions). We test headlines, content titles, email subject lines, social posts, images… anything we suspect can give us additional lift.

Often the tests don’t produce massive uplift. But sometimes, they really do – and it’s not always the result you’d expect (testing is a humbling experience for marketers who believe in accuracy of their instincts!).

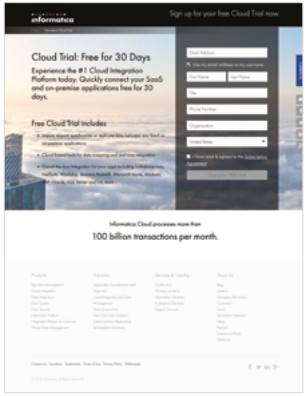
Here is an example A/B test of different designs for our Cloud Trial registration experience.

↓
fig. 24
The results show B as the clear winner at a confidence interval of 94 percent. The resulting lift is 13 percent meaning that for every 100 people viewing the page, we're getting four more to register.

A



B



	Experience	Visitors	Conversion Rate	Lift	Confidence
	Campaign	100.00% (2,053)	31.22% (641)	---	---
	Default - ION form [CONTROL]	49.25% (1,011)	29.28% (296)	---	---
	AEM trial page push winner	50.75% (1,042)	33.11% (345)	13.09%	

Start with personas

Like most marketing teams, we started our content marketing by saying, “We need top-of-funnel content for Product X,” or, “We need bottom-of-funnel content for Product Z.”

That’s fine to get started but it hits the wall pret quickly. As your matrix gets more granular – product x industry x purchase stage x persona – the content production can’t keep up.

A product-driven content program also misses an opportunity to be maximally helpful to your prospects. After all, you might have three products that would be great for a given prospect. Do they need three different pieces of content to discover that?

Today, we're building our content program by starting with the prospect – instead of with the product. We create detailed persona documents that capture each prospect type, think about the problems they have and produce content to help them. This content may ultimately map back to several relevant solutions but that comes later. First, we help people answer their most pressing questions.

A persona-driven approach involves thinking more like a publisher – putting the reader first – and less like a hard-sell salesperson.

Collect data

Traditional content is one-way. The vendor writes it and the audience reads it.

We're starting to deploy interactive content that asks the user to answer questions about his or her situation or challenges or technologies – then connects that data to their profile in the Data Lake, so we can improve the relevance of our marketing.

This approach feels really promising: we can add extra value by providing personalized reports and recommendations to each user. And we can learn more about each user to guide future interactions. Win-win.

Follow the data

Our Data Lake makes it possible to track the effectiveness of our content with a degree of granularity and accuracy that was never before possible. Because we include content tags in the marketing program codes, we can see how it performs against every metric that matters.

That means we can keep tuning our content programs as we go. Seeing which kinds of content generates the most new names; or contributes the most revenue; or leads to the best form fills. (See the Big Beautiful Bubble Chart chapter)

Content marketing plus data-driven orientation is an unbeatable combination.


Q
&
A

Q&A with Doug Kessler

Co-founder of Velocity Partners

Together with his partner Stan Woods, Doug Kessler leads London- and New York-based Velocity Partners, one of the most dynamic content marketing agencies in B2B. He's worked closely with our team at Informatica since the beginning of our big data marketing journey. We asked him a few questions about content, creativity and the power of data for content marketers:

Q What makes the best content work?

Doug I wish there were a magic formula that determines why some pieces leap out of the pack and others trail along behind.

If we look at 'home run' content across a wide range of B2B clients, it's a hugely diverse range. But a common theme seems to be a super-relevant, timely issue, treated in a fresh, confident way that's designed to help the reader as much as it possibly can. Generosity of spirit and real utility are big drivers. People respond to that.

Q Does the data-driven approach constrain content creativity?

Doug Absolutely not. Data is enormously powerful in showing us what content is likely to resonate and what content already does resonate with a given audience. It gives solidity to any brief.

Data is the starting point. We still need to come up with great ideas and craft compelling content to deliver on those ideas. We love a good hunch, but we'll take data any day.

Q How does content marketing change in a Data Lake environment?

Doug In two main ways: 1) There's just far more insight to guide our content programs. And 2) We can better justify content marketing in general and the most successful programs in particular.

With the Data Lake dashboard open in our planning meetings, we waste a lot less time on stupid marketing debates about things like whether that carousel on the Resources Page works or if that video is better than that ebook .

Fortunately, the data has moved us to create more and better content! If it indicated that banner ads are the future, it would wipe that smug smile off my face.

Q Where is content marketing going?

Doug We're seeing more interactive content, more personalized content and more content that breaks free from the purely text-based media. Some exciting formats are emerging that tell stories in a natively digital way, and maybe collect data as they unfold.

It's a fantastic time for any marketer who loves to learn.

Not so great for arrogant marketers who already know all the answers.

“By visualizing information, we turn it into a landscape that you can explore with your eyes.”
David McCandless,
Information is Beautiful

The Big, Beautiful Bubble Chart (and our new attribution model)

The beauty of a Marketing Data Lake is that whatever you decide you really need to see, there's a way to query the data, create a report and find a clear, simple way to visualize it.

This chapter is about a great new chart we've created that does things we've never seen before: our own version of the Campaign Opportunity Influence Analyzer, also known as the Big Beautiful Bubble Chart.

More than a pretty face

Like every truly great graphic, this one is more than a shiny visualization. It's actually the pretty face of an underlying analytical framework that lets us see how our different marketing content, channels and campaigns ('programs' in Marketo terms) perform relative to each other.

The idea is to clearly visualize the answer to the question, "What

works?" and to do it in such a way that all of our marketing stakeholders – product, field and corporate marketing in this case – can be in the same discussion around the same metrics.

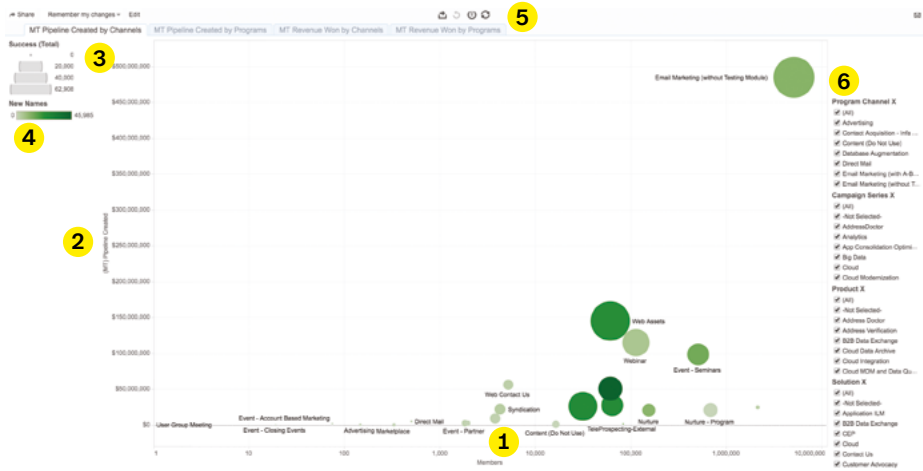
Why we needed to do this

Marketo already does their own version of an Opportunity Influence Analyzer but it didn't do what we needed. Theirs is a nice way to visualize the touchpoints on any given prospect's journey as they accumulate engagement points and become increasingly sales-ready.

But we found the Marketo approach a bit unwieldy. Marketo is dependent on costs timeframes. Earned programs that have no cost (like web visits) aren't easily included. So we designed our own that suits our business needs. Here are some of the questions we wanted to find answers for:

- Which campaigns or channels had the highest impact on opportunity creation and closure in Salesforce?
- What's the ROI for paid campaigns such as Search, Display and Remarketing?
- What campaigns or programs are the best at new names acquisition?
- Can we report on campaigns/channel performance by different dimensions such as time, region, themes, products etc.?

Those are a lot of important questions to ask of one chart. But the Big Beautiful Bubble Chart rises to the challenge.



The joy of bubbles

↓
fig. 26
Tableau dashboard showing marketing influence on opportunity creation by channel.

Above (fig. 26) is a wide view of the bubble chart.

It may not look like much but after a brief tour, we think you'll see why we love it (and why you should have one to guide your own marketing).

Here's how it works:

1. **The X Axis: Campaign Members** – the number of prospects included in the given campaign, program or tactic.

2. **The Y Axis: Pipeline Created** – the value of the opportunities (open opportunities only) that each campaign touched.

3. **Size of the bubbles** – the degree of success for each campaign (bigger is more successful, or more actual wins).

4. **Color of the bubbles** – how successful each tactic is in acquiring new names (darker means more successful).

5 **The Tabs** – this chart is actually a lot of charts; the tabs let you switch from the Pipeline by Channels view to, say, the Revenue Won by Channels – or either of these by program (specific campaign) instead of by channel.

6 **The Filters** – a powerful feature that lets us slice our data any way we like. So we can look at different timeframes; the influence of different tactics by product or solution; by specific medium or tactic; by type of campaign, etc.

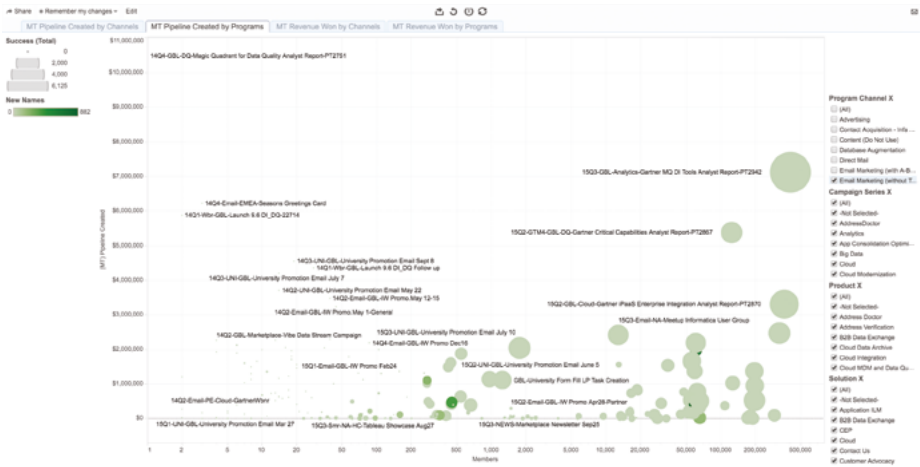
The Channel View

The above chart shows the aggregated view of all our data by channel: email, webinars, events, etc. Looking at aggregated data like this is great for comparisons and indicators – it doesn't have to be perfect to the third decimal point to deliver a lot of value.

In the Channel View, we can see that one channel, email, leaps out of the pack for dollar value of multi-touch pipeline created. We would expect this as we run more email programs than any other tactic.

But bubble size and color tell a bigger story. The chart shows us that web assets (gated content on our site) and paid media (PPC) are critical for acquiring new names (bubbles are a darker green) and are just as powerful for driving actual successes as email is (size of the bubble).

This is where the power of a clear, multi-dimension visualization comes into play. It would be easy to over-index our budget on email if we only looked at a simple metric, like pipeline created. The added nuance of success-driving and new name acquisition tell a different story – and we see them all together in one view to make an informed decision on marketing mix.



Drilling down: by program

The bubble chart is the data visualization that just keeps giving. Fig. 27 is a drill down: Multi-Touch Pipeline Created by Campaigns (or Programs in the Marketo parlance).

The snapshot here appears almost random, with no discernible pattern. It is for email programs. It's hard to see that any program is working better than any other.

Why? Because email is swamping out every other program. We've got almost half a million members opted in to our email program, while, for comparison, webinars only have sixty thousand.

So we created another Pipeline by Channel view that excluded email (fig. 28).

All of a sudden, some clear winners emerge. Web Assets (content) and Webinars leap out of the pack. Content is the winner in all metrics, pipeline created, members, successes, acquisitions and new names (Velocity, our content marketing agency, loves this view.)

↓
fig. 27
Tableau dashboard showing marketing influence on opportunity creation by programs (email as a channel).

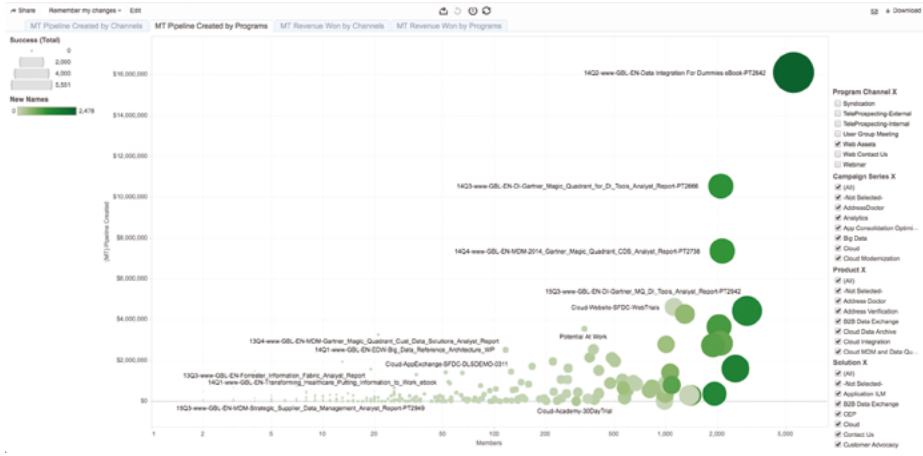
↓
fig. 28
Tablaeu dashboard shows marketing influence on opportunity creation by channels (Removed Email Marketing as a channel).

By the way, one exercise that makes the big bubble chart very actionable is to look at this view by different product lines (or countries). By comparing the same view for different products you can immediately see if the marketing mix (use of channels) is different and if that's a good thing or adjustments are required (e.g. a marketing channel not being used or overused, or a channel not working as well for one product vs. another).

Drill down again: specific content

So we know that content is our killer program. But which content performs best? Click a filter and we can see (fig. 29).

Way off in the upper right is our Data Integration For Dummies book. It's killing it on every metric. Next best are some of our 3rd-party, research-driven content assets like the Gartner Magic Quadrants. But we can also see lots of home-grown content winners here.



An important caveat: promotion

↓
fig. 29
Tableau dashboard showing marketing influence on opportunity creation by programs (website as a channel).

One thing that the Big, Beautiful Bubble Chart does not yet include is a sense of how much promotion was put behind each tactic, program or piece of content. The same is true for the length in the market – be careful comparing the impact of content offers that have been in the market for only a few days against those that have been in market the entire quarter or year.

This is, of course, a critical input in any ROI or budget optimization discussion. For instance, the For Dummies book was featured in display ads, retargeting, social media and on our home page. So you'd expect it to do well.

Right now, we manually compare the bubble chart with our paid programs to analyze ROI. But we're looking into a way to include the promo spend in the bubble chart so we can control this important variable.

A new attribution model

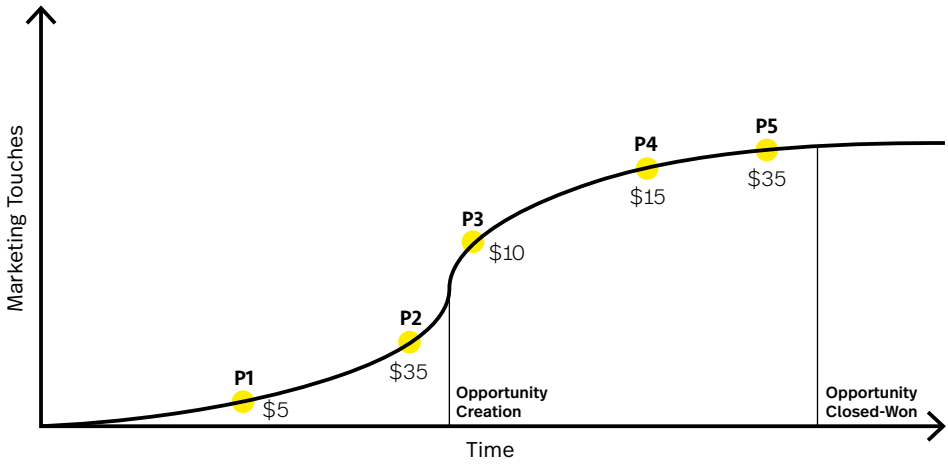
Attribution Modeling is an important tool to help you figure out which of the many marketing touchpoints in your customer journeys are making the most impact. But with so many touchpoints in a typical B2B purchase journey, accurate attribution is quite a tricky challenge.

In Marketo's Influence Analyzer, each touch on the way to becoming an Opportunity is given equal weighting (linear attribution). The most common approach is to give all the credit to the last touch, the thing the prospect did just before becoming an Opportunity. Still others use first-touch attribution that gives all of the credit to the thing the prospect did that got them into the funnel.

All three of these felt far too arbitrary for the long, complex, multi-touch buying journey that most of our customers go through. We plugged each one into the analytical framework that drives the bubble chart and we just didn't feel good about what we saw.

So we sat down with our data scientist and came up with something better. It's called the Blended Time-Decay and Positions Model and here's how it works:

> **The 'Position' part of the model allows you to give specific touches (positions in the journey) more credit than others.** We decided to credit 35 percent of the opportunity value to the program to the led to the creation of the opportunity. Similarly we gave higher value to the last touch before the opportunity for 'getting someone over the line'.
>
> **The 'Time Decay' part gives decreasing credit to all other interactions (not weighted positions in the journey) as they go further back in time.** So an email open that happened last week might get 15 points while an email open from 3 months ago might only get 10 points (shown as dollars in fig. 30).



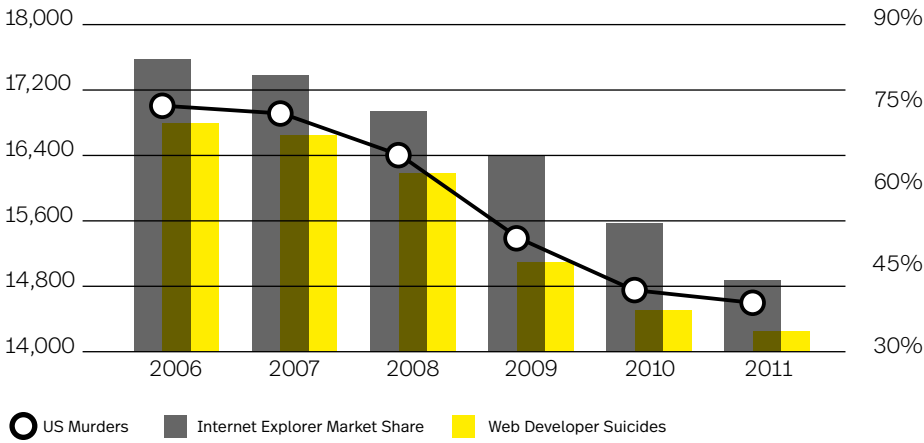
↓
fig. 30
$100 worth of opportunity that was Closed-Won. Five programs (P1 to P5) influenced opportunity creation and closure.

Is this model perfect? Of course not. But we feel it represents our actual buying journeys a lot better than the equal-weight linear, first-touch or last-touch models out there. You get the idea: if you're in B2B, you need a way to give credit to the many touchpoints that your prospects experience on their way to a purchase.

Any such model will be built on assumptions. You just need to make sure that your assumptions are consciously made and that they're exposed to whoever will be consuming the analyses driven by your model.

A survival model

As we got really excited about a more sensible way of judging the value of the different offers and tactics, we felt compelled to take another step and climb the Everest of attribution modeling – move from correlation to causation. Just because two things are correlated doesn't mean that one is a cause of the other. Take a look (fig.31.).



After some research, our Data Scientist zeroed in on a survival model used in the pharmaceuticals industry to establish a causal relationship between drugs given and survival rates of the patients. As long as you have a large enough pool of events (e.g. patients given a different mix of drugs along with their survival data), you can use these algorithms to determine a probability for a relationship to be causal and how big the lift on survival or death rate is.

When we applied this kind of model to opportunities surviving infinitely (not turning into revenue) vs. the patient 'dying' (ironically this is our desired event: an opportunity closed and won), we could see the offers, programs and channels that have true and statistically proven impact on revenue.

Here's a quick example (we're still working on making this Big and Beautiful, see fig. 32).

The power of a picture

The Big Beautiful Bubble Chart and the attribution model that drives it have become indispensible to our marketing teams:

We've replaced a vague sense of what works with a clear, data-driven view.

↓
fig. 31
It's probably safe to say that the US murder decline and the perfectly correlated Internet Explorer market share decline have nothing to do with each other. The suicide rate of Web Developers may need more study.

Data from www.reddit.com/r/geek/comments/173dcn/internet_explorer_vs_murder_rate

View By

Program Offer

Attribution

Level	Won Cloud New + Upsell ACV USD (Credit)	Won Opportunities Credit
Trial	$10,478,595	230.06
Analyst Report	$3,711,231	70.92
Other	$3,530,972	75.52
Ebook	$2,448,596	47.55
Contact Us	$1,764,620	37.79
Webinar	$1,707,054	29.74
Whitepaper	$1,462,756	30.51
Trade Show	$1,220,404	28.45
Demo	$570,941	10.84
Multiple Offers	$535,817	10.32
NULL	$522,234	8.38
Seminars	$434,318	7.85
Slide Share	$284,303	7.26
How-To Guide	$116,879	3.18
Product Update	$7,642	0.28
User Group	$6,516	0.07
Brief	$5,304	0.17
Customer Success Story	$2,280	0.03
Data Sheet	$2,259	0.05
Video	$358	0.02
Case Study	$152	0.00
Infographic	$72	0.00
Virtual Event	$10	0.00
Null		

Attribution Bar Chart

↓

fig. 32
A surival-type model for our Cloud product portfolio.

– We're putting more budget into things that work and less into things that don't.

– So we're generating more opportunities without spending more money.

– We can defend our investments in our marketing programs and argue for more money from a position of strength.

It's another example of the power of treating your data as a strategic business asset instead of just as the by-product of marketing. Get this right first and there's nothing you can't do.

“The Revenue Cycle Modeler is the central nervous system of Marketo.”
Mark Emond,
DemandSpring

The Revenue Cycle Model

A Revenue Cycle Model (a Marketo concept) is a picture of the stages your prospects go through as they progress into opportunities and, ultimately, closed deals. To us, it's an absolutely essential part of any modern, data-driven marketing department. It doesn't have to be Marketo's model, but having definitions for funnel stages, engagement, lead and sales stages etc. is a must have.

Designing and executing your model gives your entire sales and marketing team real visibility into the state of your revenue pipeline. You can see which stages are full of potential, which need topping up and which are causing opportunities to stall.

With visibility comes control. Today, we can see at the beginning of a quarter how the quarter is likely to close. If there are any concerns, such as a dip in form fills for given product line, we know about it in time to take action.

In this chapter, we'll take you on a tour of the Revenue Cycle Model that we use today.

A new model

Just three years ago, our sales and marketing model was pretty simplistic: we gated almost all of our content and we called everyone who filled out a form on our website.

That didn't just waste the time of our Sales Development Reps (and it did waste a lot of it), it also annoyed our prospective customers and caused us to overlook the opportunities that really were worth pursuing. It also generated a nominal 'pipeline' figure that never really converted into revenue – an unsustainable state of affairs.

Today, we've ungated all but the chunkiest pieces of content and let our prospects graze much more freely. We also follow a fairly simple but quite rigorous revenue generation model that we've developed iteratively, working with our colleagues in Sales. And it works much, much better.

We see a lot of marketing teams that are form-crazy. They'll do anything to capture that name. But in our experience, this is short-sighted. Letting prospects consume more content before they opt to fill out a form has ultimately led to more form fills, a healthier pipeline and more revenue at the end of the day.

In this chapter we'll take you through the Revenue Cycle Model we use today and an indicative Funnel Performance waterfall to show you how it works.

A hybrid approach: ABM plus lead nurturing

Although a lot of our big data marketing program takes an account-based view, our revenue model is oriented around the specific lead. This hybrid approach gives us the best of both worlds: a view of the behavior of entire buying teams within accounts; and a way to trigger specific actions against individual prospects.

We saw the account-based view – an invaluable resource for our

sales teams – in the Dashboard chapter. This chapter is about how individual leads turn into revenue.

Why we don't do lead scoring in Marketo

Traditional lead scoring is a powerful technique but it works best in relatively simple, single-product companies. Informatica is a complex company with many products, lots of different programs and defined sales territories.

We did start with basic lead scoring (whitepaper +20 points, Webinar +30 points, visit to jobs page -40 points, etc.) but found that it was throwing the wrong prospects over the wall to Sales. So now we use a combination of Predictive Lead Scoring, using Lattice Engines (discussed earlier) and rule-driven Task Creation that triggers sales activity differently for different programs. More on this in a bit.

The Revenue Cycle Model

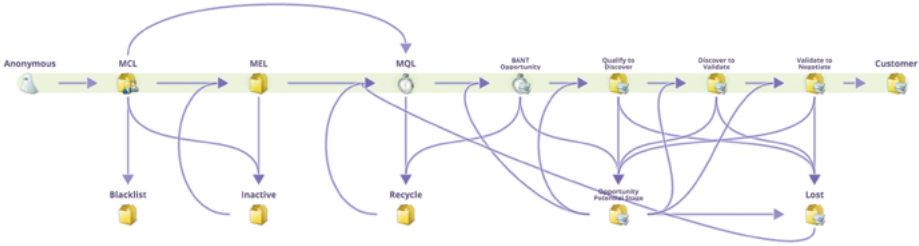
Today, we use Marketo's Revenue Cycle Modeling tool because that's where we do our email-based lead nurturing. We'll talk later about the limitations of this, but for now, it's far better than the 'pursue everything' approach it replaced.

Today, Marketo comes with a generic revenue model, out of the box. It shows six stages along a green 'Success Path' (fig. 33).

↓
fig. 33
Marketo: Analytics tab → Marketo Examples.

The Success Path is what a prospect follows if they move from stage to stage with no pause or interruption and no lost deals (wouldn't it be great if life worked like that?). In the model, a lead can only be in one bucket at a time.

Our own Revenue Cycle Model is a bit more complex but based on the same paradigm (fig. 34).



↓
fig. 34
Marketo:Analytics tab → Revenue Cycle Model.

Our Success Path has nine stages (along the green, 'money' line), with some Detour Stages along the way (the five buckets below the green line). The purple arrows show you all the ways that a prospect can move between stages.

Let's walk through it, from left to right – from an unknown prospect to someone who's paying us money:

Anonymous
This bucket includes everyone who interacts with us but hasn't yet made themselves known (by filling out a form or being manually uploaded in a list).

All our efforts here are designed to get them to interact more with us – but especially to fill out a form so we can move them along towards our sales people.

Marketing Captured Lead – MCL
As soon as an Anonymous prospect can be associated with a name (an email ID), an MCL is created in Marketo and their data is synchronized to salesforce.com (since we are using full bi-directional sync).

→ **Detour:**
Blacklist – Informatica employees and anyone from a list of competitors is grabbed out of the Success Path so we don't market and sell to them and distort our numbers.

Marketing Engaged Lead – MEL
When an MCL fills out a form, visits a certain number of web pages or clicks on certain email links, they become a Marketing Engaged Lead.

A prospect can spend weeks as an MCL before moving up to MEL status – or they can jump from MCL to MEL to MQL in a few seconds (if they fill out a high value form).

→ **Detour:**
Inactive – Leads that fail to progress from the MCL or MEL stages in 90 days are moved into the Inactive bucket. We then run programs to try to reactivate these people but some never budge.

Marketing Qualified Lead – MQL
Filling out a high-value form will trigger the move of a prospect from the MEL to the MQL stage. This creates a Task in Salesforce for follow-up by a Sales Development Rep – our handover point.

BANT Opportunity Stage 0
Based on the initial conversation with the SDR, the lead is converted into Opportunity in Salesforce and data about Budget, Authority, Needs and Timeline are captured.

→ **Detour:**
Recycle – If the lead is stuck in MQL for 90 days, or the SDR determines that the prospect is not ready to progress further, they send them back – recycle them – for further nurturing (inclusion in email and marketing programs).

Opportunity Stages: Qualify-to-Discover; Discover-to-Validate; Validate-to-Negotiate
These next three stages are owned by the sales team, who use their own definitions and trigger points to determine when a prospect moves from one to the next.

- **Detour:**
  Opportunity Potential – Opportunities are moved into this Detour bucket if they need more interactions with Sales before qualifying to move on.

Customer
When the sales team has succeeded, the prospect actually buys something and becomes a Customer.

- **Detour:**
  Lost – If we lose the deal to a competitor, the ultimate Detour, the prospect is moved into Lost bucket – until they move themselves back to the MQL stage by showing interest in another product.

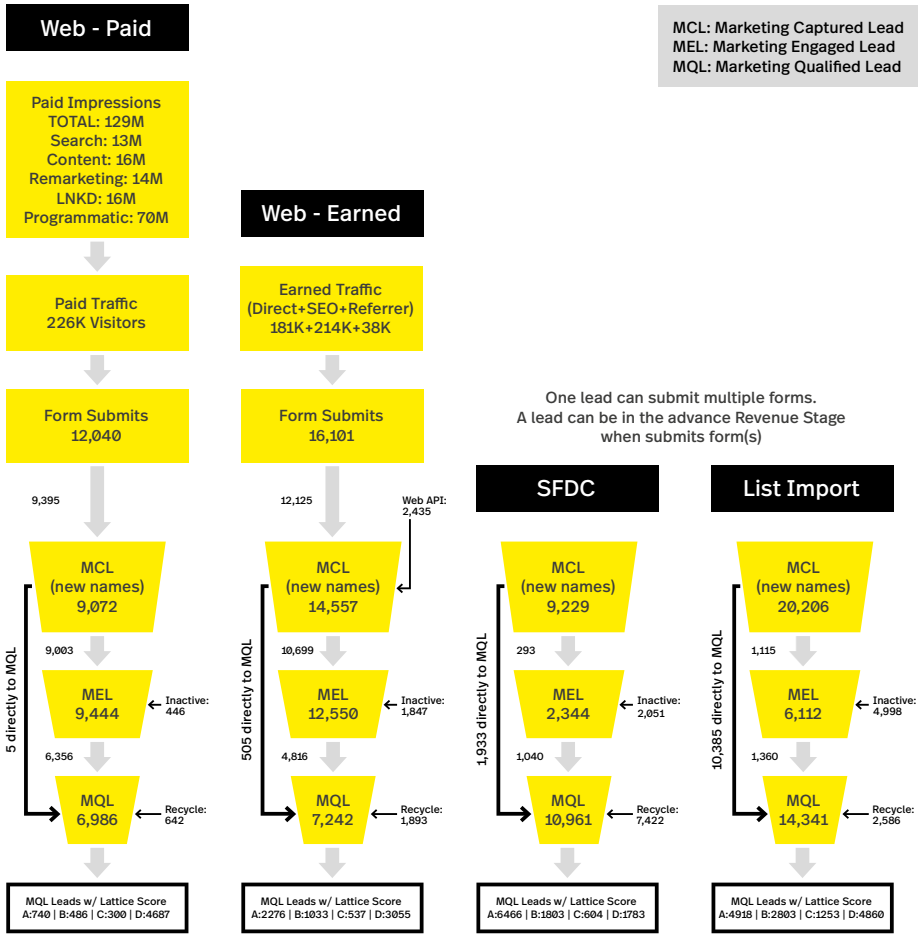
Perfection is the enemy

Revenue Models like this are a combination of art and science. The idea is not to be perfect, it's to give you a good sense of the state of your sales and marketing pipeline at any given time.

We could spend months and months trying to hone the model before releasing it. But the best way to improve the model is to use it, daily, in collaboration with the sales team.

The Top of Funnel Performance chart

To do that, we created another, custom visualization using Marketo data. Here's an example, generated for a random slice of time (fig. 35).



↓

fig. 35
Funnel report showing funnel performance by different acquisition source (data comes from Kenshoo, Adobe Analytics, Marketo and Lattice Score).

The Funnel Performance chart is a waterfall model of the first stages of the Revenue Model (through MQL), with actual numbers in each bucket, so you can see how many people are in each stage.

You can generate a single Funnel Performance chart for all of your marketing but we've divided ours into four streams, each with its own dynamics:

Web Paid – prospects that come in through paid programs like pay-per-click, LinkedIn ads and other paid social programs.

- → You can see that our Web Paid programs generate a similar number of MQLs as our Web Earned programs – but the Lattice scores for Earned are far better (see the boxes at the bottom of each column – 'A' is the best score).

Web Earned – prospects that visit our website directly (or via organic search).

- → A great channel for new names and for higher Lattice scores.

Salesforce (SFDC) – prospects that were created by Sales and brought into Marketo for nurturing.

- → Not surprisingly, these prospects have the highest proportion of Lattice 'A' scores.

List Import – prospects imported as a list from visits to our booths at events and conferences.

- → A great source of new names that also tend to show high Lattice scores**.**

Splitting different lead sources is important for showing the relative strengths and weaknesses of each.

Trends and questions

As a snapshot, the Funnel Performance chart may not lead to many ongoing insights beyond the relative performance of channels. But the real power kicks in when you look at trends over time.

If you can see that more and more leads are getting stuck in a given stage in a specific channel, you can start to ask questions about why – and take action to solve the problem.

The questions we find ourselves asking when we look at this report include things like:

Are our conversion rates at different stages increasing or decreasing over time?

What factors could be driving any changes in conversion rates?

Is there any seasonality to the metrics? Compare year-on-year.

Are our MQLs growing month-on-month? Are we meeting our agreements with Sales? Why or why not?

Can we add more resources to process leads and unblock critical stages?

Can we create new programs to target certain revenue stages?

Is the lead quality improving?

It's all about action

The Revenue Cycle Model and funnel reports we work with are much more than data for data's sake. Our sales and marketing teams are completely aligned around the key metrics in the critical stages – and the mutual accountability that they imply.

We look at these numbers to see what we can do better. To figure out what's working. To dial up success and minimize our failures.

As with everything else in our big data marketing program, the proof is in the data. If we're improving, our metrics go up. If we're not, we need to know about it as early as possible.

Of course, if you're going to base decisions on your data, it's critical that you have confidence in that data. And that's where metadata comes in.

“Metadata liberates us, liberates knowledge.”
David Weinberger,
Technologist, author

Managing your metadata

To big-picture CMOs, this chapter will be the most boring in the book. To marketing operations pros, it may be the most important. We hope that both will take away one idea: the critical importance of actively managing your metadata in an intelligent way.

What is marketing metadata?

Metadata is data about data. They're the tags and campaign codes and customer IDs and product numbers that let you compare data sets accurately and draw conclusions across data sources.

Clear, consistent metadata is the lubricant of the new marketing engine. If you've got it, everything becomes much easier. Reporting, new queries, trend analyses, drill-downs… robust metadata makes them all quick and easy.

But if your metadata is in bad shape, your operations will be too. The numbers just won't add up. Apples and oranges will all be

dumped into a big bucket called ‘food’ (along with pizza, foie gras and single malt). Your reports will disagree with each other. Everyone will lose confidence in your data – and the conclusions derived from it. And rightly so.

Are we overstating this for effect? No, we’re not. We’ve worked with data for long enough to know just how important it is to get your metadata act together before even attempting a big data or Data Lake project.

At the start of your big data marketing journey, managing your metadata will feel like a black art. But very quickly you’ll get a feel for it. You’ll find yourself in situations where you wished you’d made different decisions about taxonomies or data organization, and these will sensitize you to the implications of metadata strategies. Hang in there.

The implications of metadata

Metadata is there to help you look at the things the same way across different systems. Let’s look at some typical problems:

– You want to target prospects by product interest but the product IDs in your campaign management (or marketing automation) system are different from those in your ad retargeting platform.

  → Your product metadata is inconsistent.

– You’re running a campaign that targets the insurance industry but three of your systems bundle ‘insurance’ into ‘financial services’ along with ‘banking’, ‘credit cards’ and ‘capital markets’.

  → Your industry taxonomies have different grain sizes in different systems.

– You're rolling out a global software release but your email segmentation is by region instead of by country – and some countries don't even sell the product your upgrading.

  → Your geographical metadata is grouped in an inflexible way.

In these examples and many more that are common to every marketing department, **consistency of labels and of taxonomy** are the main issues.

If your email server only stores the product interest of prospects by product family, you can't target emails down to the level of individual product. And if FR is the country code for France in your CRM while in your Eloqua system it's F… you can't bring the two together without an intermediate translation step.

Important metadata for marketers

The metadata that's most important for you will reflect your own business. But most marketers will want solid, consistent metadata around the key dimensions we work with, such as:

Product – to look at important metrics like new names, engagement or revenue by product, you need solid product ID convention used across all systems and applications.

Geography – again, you can't split your analysis by state, country, region or territory if your geo codes are all over the place.

Industry – knowing you're getting traction in one industry could flag a major opportunity. Manage your industry metadata rigorously.

Marketing Channels – marketers need to know how each channel performs over time and its ongoing contribution to revenue. Make sure your channels – PPC, paid social, web,

email, etc. – each has its own tag and that these tags are used consistently everywhere.

Campaigns – tracking the success of any given program or campaign – a webinar, roadshow, event, content promotion, banner campaign, etc. – depends on consistent campaign codes that everyone uses.

Our campaign codes attach to the prospect and get passed from Analytics to Marketo to CRM, so we can see the impact of every campaign we run.

↓
Two approaches to campaign codes
– Choose intelligent campaign codes that are human-readable and share the 'recipe' with all marketers (which can get complex).

– Use a unique number (which requires a 'decoder ring' somewhere).

Your choice. We went with route one.

Some core principles

As you work with your metadata, you'll start to see how to create data structures that make sense for your business. As we did that, some core ideas started to emerge:

Try to make your metadata as consistent as possible across all of your systems. If you don't, you'll have to maintain complex mappings between systems – and that makes your data management much more complex (not impossible by any means, just harder).

Avoid taxonomies that are not based on first principles – choose simple, basic categories that won't change very much. For instance, countries don't change that much over time, but sales territories do. So if you build your geographical categories on countries, you can always roll the countries up into sales territories for analysis.

This allows you to do historical analysis simply and clearly, by creating a report that says, for instance, "What was the revenue three years ago for the sales territory we define today as Central Europe?" Even though you changed territory definitions last year, this analysis is easy because you tagged your data by country.

↓
The Lake can help
Your data mappings (showing that FR in your CRM means F in Eloqua) can be kept in a spreadsheet and loaded into your Data Lake for automated processing. You'll have to maintain the mappings over time but at least you won't have to do manual translations outside your analytics framework.

Choose a schema that's fine-grained enough to preserve meaningful distinctions. For example, in your industry taxonomy, 'Insurance' is more fine-grained than 'Financial services' but not so fine-grained that you over-complicate your life. Look for reasonable, useful granularity rather than infinite precision.

↓
You can roll ZIP codes up into metro areas for analysis but if you only stored your data by metro area, you can't later turn them into all their discrete ZIP codes.

Look for weird results. If there's a strange, sudden change in a trend or an unexpected break from what you expect from a given metric, the cause is often that a definition somewhere – the metadata – was changed back in time.

Did your Southern Region suddenly dip 30 percent between 2013 and 2014 for no good reason? Dig a bit. Aha: the region included Virginia and Maryland in 2013 but not in 2014. Luckily, you stored your data by state so you can easily correct for that. Anomaly fixed.

Beware of over-writing metadata. Often, a metadata value will get overwritten as things change. A prospect might have had a Lattice score of B when she became an opportunity, then got promoted to an A when she attended an event. That's fine and necessary. But you might want to capture a value called 'Lattice Score When Opportunity Created' to preserve the fact that she was a B when she was deemed sales-worthy.

Why? One reason would be to test the accuracy of your Lattice scores. When the prospect finally buys, you need to know that Lattice first scored her as a B and not an A.

Similarly, retaining metadata for 'Source on first entry into database' tells you that someone came from SEO even if his most recent source channel was, say a web banner ad. Your attribution model will want to know that.

Preach compliance. People can be lazy. To keep everyone using the right codes for everything you need to preach the gospel – and show them why it's important how you're using the tags.

The more you use the data – the more you make budgeting decisions based on it – the more careful people tend to get.

After all, the PPC folks don't want their budget cut just because they failed to use the right campaign code and so didn't get credit for the revenue they generated.

Start operating off the data and it gets better fast. Shy away from it and it never will.

Metadata matters

Until you start working with data, issues like metadata feel like tiny, techie process details. Once you've started to make your marketing truly data-driven, the importance of metadata just keeps rising.

Here's the thing: data doesn't really have much value at all without its metadata – the tags that tell you what the data means. But with robust, well-managed metadata, the value of your data skyrockets.

↓
In an all-hands meeting, we were reviewing Form Submits for all channels and campaigns – by product segment and marketing channel. One segment, let's say big data, looked like it had taken a major dip. One marketing team piped up and said the number didn't account for a whole lot of campaigns that didn't carry the big data tag. Guess what? Next time, they'd all been tagged perfectly.

“Map out your future, but do it in pencil.”
Jon Bon Jovi

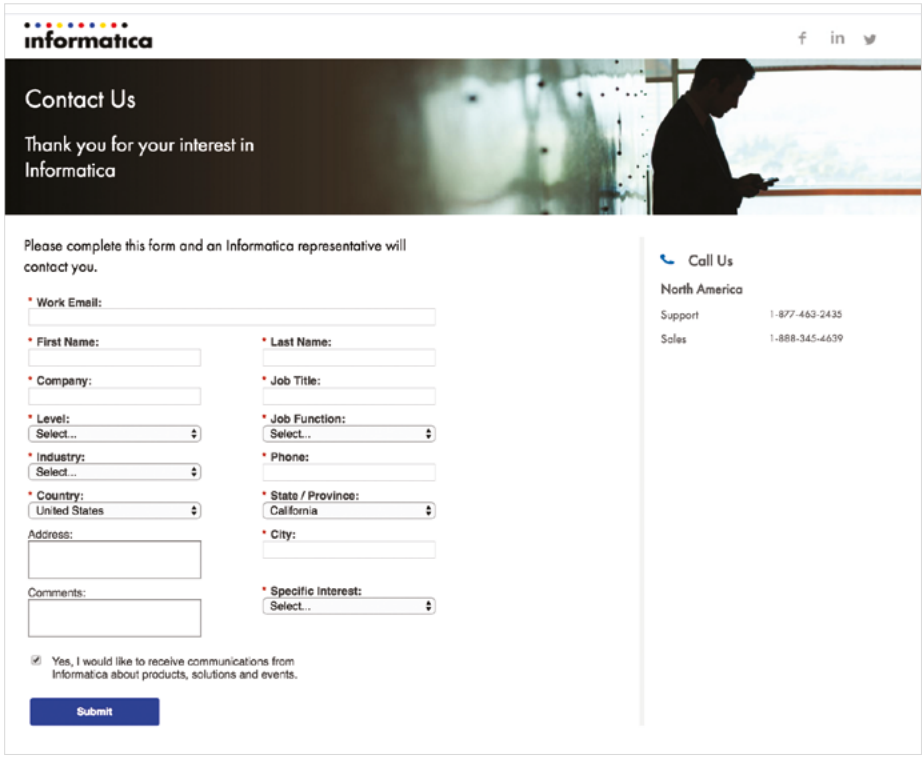
More Data Lake use cases

So far in this book, we've focused on a few main use cases: an ABM analytics dashboard and a cross-sell spotter for Sales; and the attribution and revenue models for Marketing. But the beauty of a Data Lake is that you can think up new uses and spin them up in very little time. In this chapter, we'll look at three examples.

1. 'Contact Us' form abandonment

The highest-value form on our website is the Contact Us form. It's our best indicator that a prospect has traveled through the funnel and is close to a purchase decision. And we know that the people who complete the form are far more likely to turn into revenue than people who don't.

But what about all the people who reach the Contact Us form but don't complete and submit it? Surely, that's a pretty serious intent signal that we should be acting on.



↓
fig. 36
Conatct Us form, www.informatica.com.

That got us thinking. If the Contact Us form is the closest thing we have to an ecommerce shopping basket, why don't we do what retailers do: run some 'cart abandonment' strategies (or in our case, Form Abandonment). Retailers who send emails to shoppers who have recently abandoned their carts have seen great conversion rates. Maybe we could too.

The Data Lake makes it easy to analyze this issue and take action on the results. Specifically, it helps us answer questions like these:

– How much revenue can we attribute to the Contact Us page?

– How many known names in our database have visited the page but decided not to fill out a form?

– Are visitors who do complete the form first time or repeat visitors?

– How many new names are we able to generate from this form?

The first step was to look at the form completion rate for all of the Contact Us pages across our regional sites (fig. 37).

↓
fig. 37
Adobe Analytics: Custom Conversions → Form ID (eVar32) showing Success Events Form Views (e12), Form Completions (e1) and Form Completion rate (calculated ratio e1/e12).

	Form ID (v32)	Unique Visitors		Form Views (e12)		Form Completions (e1)		Form Completion Rate
1.	marketo|us|contact us|contact us|	4,608	56.9%	4,874	52.9%	548	54.5%	11.2%
2.	marketo|go|contact us|contact us|cloud	1,170	14.4%	1,281	13.9%	154	15.3%	12.0%
3.	marketo|in|contact us|contact us|	644	8.0%	662	7.2%	47	4.7%	7.1%
4.	marketo|go|contact us|contact us|marketplace	360	4.4%	368	4.0%	52	5.2%	14.1%
5.	marketo|de|contact us|contact us|	312	3.9%	328	3.6%	19	1.9%	5.8%

The report makes us wonder why the form completion rate for Germany is so much lower than the US rate (Cultural? Hard to say but the data highlights the issue for us).

Next, we need to look at the people who visited the page but didn't complete the form (4,100 of them for the first instance of the form). What was their engagement like? A quick look in our analytics showed that these people are highly engaged – viewing almost five pages per visit compared with two pages per visit for our average visitors (fig. 38).

↓
fig. 38
Adobe Analytics: Custom Conversion → Form ID (eVar32) showing Success Events Form Views (e12), Form Completions (e1) and Form Completion rate for people who abandoned the form (calculated ratio e1/e12).

	Form ID (v32)	Unique Visitors		Form Views (e12)		Form Completions (e1)		Form Completion Rate
1.	marketo|us|contact us|contact us|	4,079	56.9%	4,296	53.2%	0	0.0%	0.0%
2.	marketo|go|contact us|contact us|cloud	1,001	14.0%	1,091	13.5%	0	0.0%	0.0%
3.	marketo|in|contact us|contact us|	595	8.3%	612	7.6%	0	0.0%	0.0%
4.	marketo|go|contact us|contact us|marketplace	313	4.4%	315	3.9%	0	0.0%	0.0%
5.	marketo|de|contact us|contact us|	291	4.1%	306	3.8%	0	0.0%	0.0%

	Date	Page Views/Visit	Visits/Visitor
1.	Dec 2015	4.74	1.06
2.	Jan 2016	5.09	1.06
3.	Feb 2016	5.02	1.07

Now let's move into Marketo. Whenever someone fills out a Contact Us form, their data goes into Marketo, then on into

↓
fig. 39
Engagement for the segment that abandoned the form.

Salesforce. As you can see in fig. 40, 806 people became members of the Contact Us program and 39 percent were new names. Clearly, the form plays a big role in new name acquisition.

↓
fig. 40
Marketo RCE: Program Membership Report.

Program Chart	Members	New Names	% New Names
NA-Ongoing-15Q1-Contact Us Page	806	312	38.71%
Grand Total	**806**	**312**	**38.71%**

For the same time frame, the Contact Us form led to $9.7m in pipeline creation in Salesforce. That means that each form completion is worth $12,000 to us.

Program Chart	Program Name	(MT) Pipeline Created
Web Contact Us	NA-Ongoing-15Q1-Contact Us Page	$9,681,462

↓
fig. 41
Marketo RCE: Program Opportunity Report.

So if we could convert 10 percent of the 4,100 visitors who abandoned the Contact Us form on the .com site alone, it could lead to an additional $5m in pipeline. Does that sound like an opportunity worth pursuing?

So what might we do to achieve that 10 percent uplift? Here are a few ideas:

- Run an A/B test to optimize the form.
- Offer a chat session to visitors who abandon the form.
- Conduct a qualitative survey to find out why people abandon.
- Send relevant emails or make SDR calls to the people who abandon the form (it's now exposed in the ABM dashboard).

So that's the Form Abandonment use case: drilling down into the data to spot a revenue opportunity, then taking steps to realize that opportunity. Having the ability to dig into data, find opportunities and size them up, then action a fix and measure the results, provides instant gratification. Anyone seeing the ROI on data yet?

2. Voice of the Customer analysis

The next example is about adding a new data source to the Data Lake and seeing what insight it generates. In this case, the new data source is the results of a customer satisfaction survey.

It's a simple idea: run a pop-up survey to a slice of our web traffic, as they exit the site. Ask a range of customer satisfaction questions about the purpose of their visit, the site experience and the user's general experience with (and perception of) us. Then marry this data with our clickstream and marketing program data in the Data Lake.

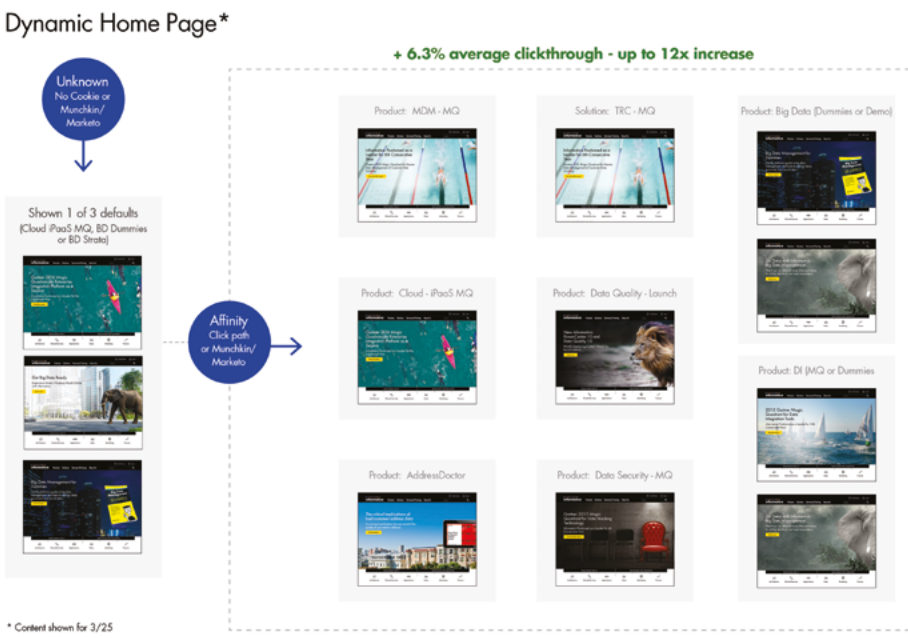
Starting with the CSAT score segments, we can start asking questions like:

- How do key dimensions and metrics – engagement level, funnel stage, etc. – vary by CSAT score?
- What site sections or pages correlate with high or low CSAT scores?
- What click paths lead to better or worse satisfaction scores?
- Is CSAT different for known vs anonymous visitors? For different industries or segments?

As you can see, simply adding a new source of meaningful data opens up a new world of analysis, allowing you to run the new data against all existing dimensions and metrics. It's a great example of the open-ended agility of the Data Lake concept.

3. Segmentation and personalization

The days of one-size-fits-all marketing are coming to an end. But it can be quite hard for a marketing department to kick the habit of sending everything to everyone.



↓
fig. 42
Our generic home page has about 1 percent click-through rate (CTR). Our targeted home page sees an average CTR of 6.3 percent, with some pages achieving as high as a 12 percent CTR.

The Data Lake helps us split our target audience into meaningful segments based on a wide range of relevant dimensions. It also lets us trigger specific actions for specific prospects: the segment of one. Finally, it lets us track the results of our efforts to see which segments and which tactics generate the most returns.

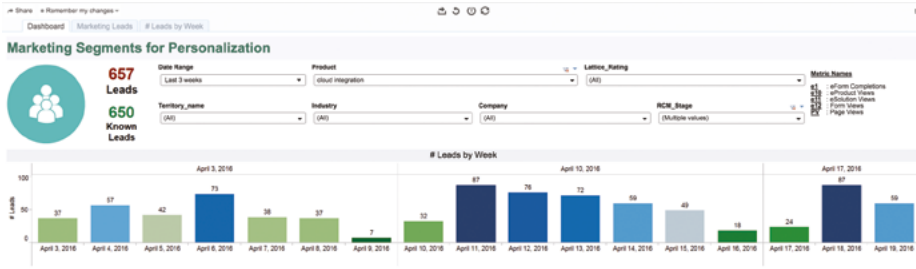
Segmentation and Personalization are major use cases for us. We've seen such great results from personalizing our web pages, email programs and paid media campaigns, that we've made improved segmentation and targeting a key priority. The Informatica.com home page is a great example for what happens when you listen to your customer's behavior (fig. 42).

The segmentation dimensions we started with include:

Web behavior – visit recency; product interest, etc.
Firmographics and industry – from Demandbase data
Territory – from Salesforce

Revenue Cycle Stage – from Marketo
Propensity to buy – from Lattice Engines

Using these dimensions as filters, allows us to see segment sizes along with the email IDs and clickstream data for members of each segment (fig. 43).



↓
fig. 43
Tableau report showing leads that have shown interest in Cloud Integration product for the past three weeks.

With this information, we can target with as fine a grain as we like, delivering targeted emails from Marketo, or custom web content using Adobe Test & Target. We can also deliver our segments to our third party providers of paid media programs, to target in social media or in banner ads.

Our segments become tags that stay with prospects throughout their journeys, so we can see the impact of our segmentation strategies as they trickle through to revenue – and see the uplift versus the non-segmented audience.

The end game: a consistent, cross-channel customer experience that lets us deliver more relevant, more timely messages (and eliminate the irrelevant stuff so many marketers bombard people with). Optimizing within any channel is good. Optimizing across channels is where we're headed.

Again, these are early explorations for us, but we predict that this use case will become one of our most compelling Data Lake stories.

The gift that keeps on giving

The examples we just summarized show the enormous power of a data-driven approach to marketing, when powered by the flexibility and extensibility of a Data Lake.

New ideas, hypotheses, strategies and tactics are all made possible and accelerated by a single, central, connected data repository based on a schema-on-read paradigm. And because the data is right there, you can model your ideas to evaluate their potential before you put resources into execution.

We anticipate many more use cases for our Marketing Data Lake. And each one will increase our ability to generate new revenue for the business – and prove it with the data.

“If you’re not making mistakes, then you’re not making decisions.”
Catherine Cook,
Co-founder of
myYearbook

Implementing your Data Lake program

We've gone through the basic principles of the Marketing Data Lake and its use cases and drilled down into some of the operational issues relating to integrating different data sources.

So how do you get there from here? Where do you start and how do you proceed? That's what this chapter is all about.

Driving – and managing – change

The first – and possibly most important – thing to realize when starting out on a transformation to a big data operation with a Data Lake at its heart is that this involves a significant culture change. The more you see it as a change management process that needs active stewardship, the more successful you'll be.

It's not enough to know that you're doing the right things, you have to bring the organization along with you. That means listening hard,

communicating clearly, picking your battles well and marketing your successes.

Here's a summary of the essential steps in leading your transformation.

1. Establish your baseline

The first step in any transformation is to understand where you are today. In a Data Lake program, that means getting answers to these kinds of questions:

Your Data
What data do you have?
Where does it live?
How good is your data? How clean? How complete?
How old is the data? Has it been validated or verified recently?
Is your data segmented anywhere? How do the segments perform?

Your Funnel Dynamics
What are your conversion rates at each stage?
How many leads turn into opportunities?
How many opportunities turn into revenue?
How many form fills do you get every month? Where?
How many emails do you send?
What are your open and click-through rates?

Your Processes
Who manages metadata: campaign codes and tags?
Is any data governance in place?
How are leads handed over to Sales?
How do Sales and Marketing work together?
Are there any SLAs between the two teams?

The answer to many of these questions will be, "We don't know." That's okay. The first step is to know what you don't know – then to take steps to find out.

↓
Find some benchmarks
As you expose your key metrics, the natural question will be, "Is that a good number or a bad number?" There are lots and lots of sources on the web for benchmarking your key metrics. Even if they're not exactly from your industry or market, they're a start. Get Googling.

Here's an example of published benchmark data from Insight Venture Partners (fig. 44):

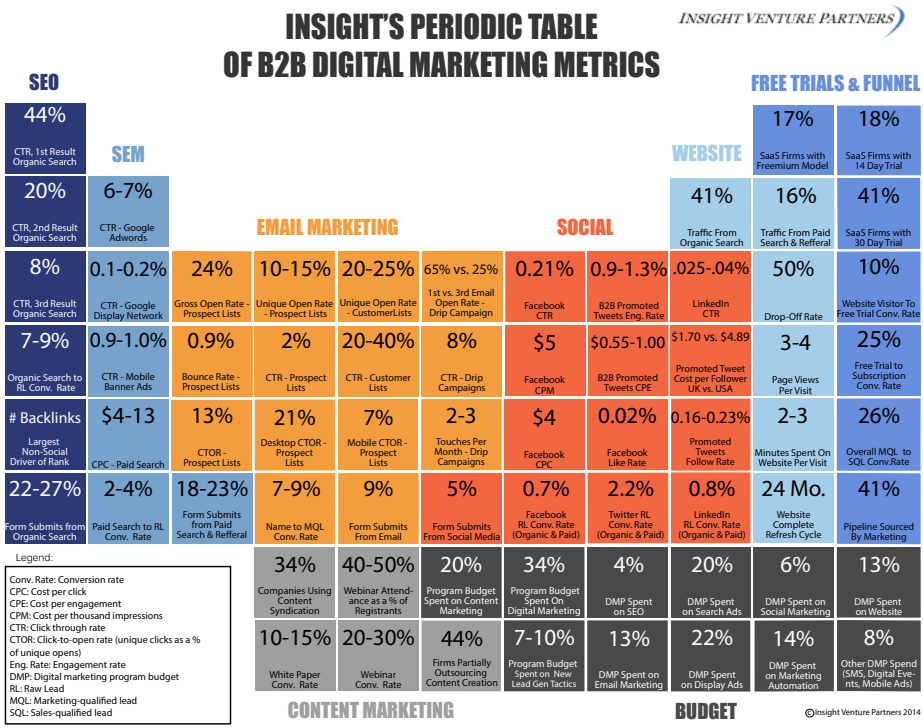


fig. 44

2. Get an executive sponsor

When you connect the dots across different marketing channels, you'll end up moving budget between them. The people who lose budget won't be pleased about that – unless they can see the bigger picture. A CMO or Head of Corporate Marketing can help them see the bigger picture. So get one on board early.

You might have expected this to be step 1 but we actually think it's easier to get a great sponsor if you know how much pain you're in first. Either way, you'll need air cover for an initiative of this scope and scale.

We were lucky: our executive sponsor happened to be the leader of the project (Franz). He already had the backing of the CEO, CMO and board. If that weren't the case, we'd have spent as much time as it took to find that sponsor, explain the journey and get buy-in.

Why? Because every project like this is bound to hit some rocks. And nothing moves rocks better than someone with a C in their title who really understands the value of the trip.

3. Fix the worst problems

Chances are, you're going to expose some nasty truths in step 1. Some of these won't be crippling but others will. So before you do anything else, figure out how you'll fix your biggest problems.

A good place to start is to clean up the implementations of your three main marketing applications: analytics, marketing automation and CRM. In some cases, you may need to start over, with a clean deployment or by migrating to a new application. Yes, migrations of such critical applications are painful. But they're also a great way to sweep out a whole lot of legacy problems in one fell swoop.

In the Five Foundations chapter, we showed you the things we had to fix before starting out with a Data Lake. You probably have similar issues. Don't sweep them under the carpet and hope they solve themselves later. They won't.

4. Pick a key stakeholder

The best way to get a big data marketing program off the ground is to find an internal customer and solve an important problem for them. The first step is to understand the stakeholders you've got and to think about their perspectives and agendas. Here are four big ones:

Sales teams – the people who generate revenue are your most important stakeholder. If you can unblock a problem for them, you're on the fast track.

Marketing teams – the teams behind each channel – search, social, advertising, email, field marketing – are all potential stakeholders. Improve their performance and you're on to a winner.

Product teams – the people who own products need to know what you're doing for their products and how it's working. If you can slice your data by product, you'll make friends. If you can't…

Customers – your ultimate stakeholder is the customer. If you can spot and fix customer experience issues in your program, the upside should become clear to all.

For our program, we chose to start with Sales as our first stakeholder. At heart, we're revenue guys. If what we do isn't making money, it bothers us.

5. Pick one of their problems

The next step is to sit down with your chosen stakeholders, interrogate the hell out of them and listen really hard to their answers.

What are they most frustrated about? What's broken? What would most improve their lives?

For our program, the sales people were pretty clear. The pipeline we created wasn't converting well enough. And the quality of the leads we were handing over to them were poor: Too many non-working email addresses or phone numbers. Too many form fills with Donald Duck or Bill Gates in the name field. Too many competitors or employees handed over as leads. (Note: some of the problems

you hear will be perceived problems. That's okay. Correcting a misperception can be as valuable as fixing a real problem.)

How do you choose which problem to start with as your first use case? The key is to find the sweet spot: the right combination of importance of the problem and achievability in a relatively short time frame.

For us, delivering an account-based view of prospect and customer engagements on the web and in email was a big win – and one we knew we could achieve quickly. So that's where we started.

Improving pipeline conversion was the real prize but we knew that this would take longer. So while we started on the road to better pipeline, we made the great dashboard for Sales that you saw earlier.

→ **Translating pain into pilots**

Is Marketing pipeline converting to revenue?
→ Look at conversion rate improvements

Do you know what marketing programs are working?
→ Look at attribution model and metrics

Is Marketing delivering quality leads (valid contact data, good fit, budget, authority, need, timeframe)?
→ Look at data quality, profiles, lead scores, process/SLAs

Are we losing the deals we didn't know about?
→ Look at acquisition instrumentation and measurement

Is customer experience clunky or friction-prone?
→ Look at touch point analysis and journey maps

6. Build the business case

You've got your stakeholder and your problem. Now show how much you'd make or save (or both) by solving the problem.

In our case, SDR productivity was the metric we knew we could improve with an ABM dashboard. So we built our case around that.

For the bigger play – improved pipeline – the metric we focused on was improving the conversion rate of opportunities to revenue. We showed that if we could move that rate from one to two percent, the effect would be the same as doubling our marketing spend. Now that's a business case.

Pick a number and put a stake in the ground. The data is with you.

7. Run a Proof of Concept

Before asking for the big money, you need to prove that what you're doing will work. That means a Proof of Concept (POC). The goal of the POC is to get backing to scale the use case up and get it into production. No more, no less.

Here are some good POC principles:

Limit the scope
A good POC is one where you can show results fast – but without having to wrangle and clean enormous amounts of data in order to get there. Think 'Minimum Viable POC'.

Agree on what good looks like
Establish the metrics of success and the uplift you expect.

Get it up and running quickly
We spun up servers in Amazon Web Services using a credit card, so we didn't have to wait for IT provisioning systems. Whatever works.

Build it so you can scale it
Ideally, you won't throw away your POC and start all over from zero once you get approval. So even in a dev environment, you want to use a best-practice architecture, with security and integrations that really work.

Iterate with your stakeholders
Show them what you've done and ask for feedback. Make improvements and do it again. And again. Soon, you'll have something they love.

Market your win
When you do get a happy stakeholder, you need to let people know about it. Data is key, of course, but anecdotal evidence can be really compelling too. We told a few front-line stories and showed a great demo to the board – it won us a hell of a lot of support. (Never underestimate the power of a demo.)

8. Get IT on board

A project as big as a Marketing Data Lake can't be done without the full support of the IT department. Don't even think about trying to do it 'under the radar'. If you want to use IT resources, you have to justify it.

When is the best time to get your IT folks on board? Ideally, when you've got something real to show them. But it does depend on your existing relationship with your IT people. If it's good, you can get them involved really early.

Again, we're lucky: we've got an amazing IT department and they were incredibly open and supportive of this initiative. Yes, we needed to convince them that a Data Lake was better for our purposes than the existing data warehouse, but once they saw why, they were a terrific ally.

Remember: some of this will be new to the IT department. Traditional IT has much longer cycle times and agility has not been their number one priority for many of their projects (things like security, cost and consistency were). So some of the things you're asking for will be alien.

The good thing is that IT folks do recognize the need for new ways of working – and they love new tech that really nails a problem. Our project meant learning about Data Lakes and giving business users direct access to data. They liked the sound of that and we were happy to make it happen. And ever since they've been a great partner to us.

If you haven't yet built a great relationship with IT, now is the time to invest in one. It's worth it.

9. Go into production!

The POC worked. The stakeholder is thrilled. The Board of Directors is behind you. And IT is on board. The only thing left to do: ramp up!

Now you'll have to think about things like where your data will live, who will access it, how you'll secure access… hugely important things that your partners in IT will help you with.

When we scaled up our Data Lake, we moved it from Amazon Web Services into our own hosted servers, in a virtual environment. Made sense for us and it's worked brilliantly in production.

- → **A word about QA**
  Don't underestimate the importance of Quality Assurance as you go forward with your Data Lake program. And don't think you can outsource the QA: it must be done on the business side, not the IT side.

Expect to spend some cycles figuring out why certain data looks funny (hint: something's wrong); or why those two reports don't agree; or those fields aren't mapping…

Rigorous QA means you can build and maintain trust in your data – the key to credibility. Get your hands dirty with the data. Make sure it hunts. That A wiggles with B, as expected. No shortcuts!

10. Keep the momentum

Standing up a new infrastructure and a use case or two is not the finish line, it's the starting line.

To keep the momentum up, you need to continually align your team and your stakeholders; review progress; plan next steps; and celebrate your wins. Yes that means meetings. Here are our main ones, beyond the daily cadence of dashboard discussions:

Weekly Operations meetings look at campaign management and tactical issues.

Bi-weekly staff meetings with Franz and his direct reports cover management issues, budgets and coordination across departments.

Bi-weekly all-hands meetings (alternating with the above) cover a broad view of what's going on; a drill-down into a topic (like Form Fills); a review of key metrics and forecasts; data from the Ops review; stuff like that.

Quarterly operations reviews in which we dissect a 60-80-page report over the course of a full day. This is a detailed look at the trends for our key metrics by product, country, segment… everything. Goals achieved? Why or why not? Any surprises? What do they mean?

End of year reviews – our Q4 review is even more intensive with lots of historical year-on-year analysis, soul searching, retrospection and introspection.

Go forth and deploy

When you embark on a Data Lake project, you're in the transformation business. Neglect the change management side of things and you're in for a world of pain.

But manage your deployment systematically, with the full alignment of key stakeholders, and you've more than doubled your chance of success.

“In most cases being a good boss means hiring talented people and then getting out of their way.”
Tina Fey, Bossypants

Your team and their skills

One of the first things that fellow marketers say when they see our operation is, “I don’t know if we can get there from here. If we have the people or the skills.” That’s a natural concern – and not one to be ignored.

When we look at it, the team we have today is actually very different from the team we had three years ago. Some of the people are different, there are some new roles, and the skills of the team have changed significantly.

Today, the team is far more tech-savvy, digital-driven and data-aware than the team we had even a year ago. The transformation was pretty fast and we think the whole department has really enjoyed the change (despite the inevitable growing pains).

New role

As the importance of certain channels emerged – and we shifted

our budgets towards digital marketing – we started hiring people who specialize in things like paid media, SEO and social. We also created two new roles we didn't expect:

Performance Manager
The most important new role has been Performance Manager (ours is Anish Jariwala, the co-author of this book). This has proven a critical role: someone who owns the machine and knows how to drive it.

Anish is far more than an analyst or reports guy (though he's that too). He adds value to the entire process, knows how to glue the technologies together and works with business users to drive new use cases. To do it, he has to understand the technology, the data and the way they interact with the business. In short, he enables, measures and drives our marketing performance. Without him, we'd be lost. With him we can take on the world.

So find yourself an Anish – but not OUR Anish!

Data Scientist
We didn't think we'd need a Data Scientist, but in John Teifel, we found a great one. He knows how to build data models, how to use data as a predictive tool and how to stay within statistical significance. By the way, he speaks 'R'.

With access to a data scientist, we can proceed with confidence.

New organization

The marketing operations person used to live within Sales Operations. In practice, we didn't get much of their time. So we hired our own marketing operations pro – Laura Wang – and let her build a team. The result has been dramatic. We now own the processes we need to own instead of begging for resources.

New training

We're all on a steep learning curve, so we're all always learning. With constant training programs – in Marketo, Adobe, Tableau, whatever we need. We use the vendor's training as well as designing our own in-house training programs on things like self-service data.

The goal: self-sufficiency. We encourage people to do their own things and pursue their own projects, with central assistance where needed.

T-shaped people

A lot of digital marketing departments have I-shaped people: people who are great at a single discipline. At Informatica, we like T-shaped people, who can go deep in a given discipline but also see across the piste and are comfortable with the concepts that every other team member deals with every day.

T-shaped people give us flexibility and agility. They also make each discipline better able to play nicely with others and understand the context of their decisions.

Data in our DNA

Over time, we're all becoming data geeks. We're seeing people use data to drive their decisions in new ways all the time. An example:

Running a roadshow in the old world meant picking some cities and applying a standard event formula (email invitations, follow-ups, etc.).

Running a roadshow in the new world means looking at different metro areas to spot product interest clusters; designing a program

that's relevant to the people in the area; targeting our invitations to the people who would most value the event (and leaving out people who would have zero interest).

In marketing, data changes everything. Changing our culture over time to a data-driven culture pays dividends in a hundred different ways.

“The price of light is less than the cost of darkness.”
Arthur C. Nielsen

What's next: The Customer Data Lake

So what's the future of the Marketing Data Lake? The future is beyond marketing. The whole point of a Data Lake is to collect a wide variety of data sources, integrate them and make them available for many use cases.

That means the Data Lake should be an enterprise resource, not just a marketing resource. In our view, the next stage for Data Lake is to embrace the other customer-facing disciplines, so that Sales, Marketing, Customer Service and Product teams all have access to the same, centralized resource.

Whether or not the Lake can or should go beyond the customer -facing disciplines is another issue. The Data Lake approach isn't right for some other departments, such as Finance & Accounting. But it makes sense to use the same Data Lake for all the disciplines and departments that touch the customer directly.

The silos are falling

The lines between the customer-facing disciplines are already starting to blur, and for good reason. All of them are essentially in the customer experience business, so all of them need to be able to see interactions that happen across all touchpoints, no matter where they occur.

A customer service agent is presented with sales opportunities every day. Every salesperson routinely sees service and support issues. And marketers see and can contribute to sales and service engagements (with relevant content, for instance).

A single, shared Data Lake – The Customer Data Lake – feels inevitable. So anyone on any team can see things like:

> **The entire customer journey** – including engagements with marketing, sales, service and the product.
>
> **Actual product usage** – for software, cloud services or any connected products, insight about actual usage patterns would be gold dust. In the other direction, the ability to inject sales, marketing or service messages into the product experience holds enormous potential.
>
> **Customer lifecycle insights** – linking how a customer was acquired with the dynamics of their long-term relationship. For instance, web forms may be the number one way customers are acquired – but also lead to more and quicker defections than any other acquisition source.

In short, a shared Customer Data Lake would be a target-rich environment for anyone trying to improve customer satisfaction, lifetime value, loyalty and profitability.

We're particularly excited about how the Internet of Things will bring product experience insights into the analytical framework. If your fridge can order more milk, it can also alert your ad server to deliver milk promotions for a few, high-value hours. When your

copier orders more toner, it can also schedule a maintenance visit and suggest to the customer a Guide to Printing less Color.

A unified view of the entire customer experience – powered by the Customer Data Lake – will help us all design new customer experiences delivered via new business models. The customer wins and so does the business.

“The only thing worse than being blind is having sight but no vision.”
Helen Keller

The most important fork in your career

There's never been a more exciting time to be a marketer. There's never been more change – in the way people interact, buy, sell, learn, communicate and share. There's never been more data available to help us interpret and respond to these changes. And there's never been more technology to help us manage, integrate and analyze all that data – then bring the insights to market. It's natural to feel overwhelmed by all this change. To feel swamped by the data. And bewildered by the technology. But you don't have the luxury of being overwhelmed. The business is looking to you right now to master all of this change and turn it into growth. To do that, you have to make a decision right now to do one thing: to step up and take responsibility for your data.

To refuse to keep it locked inside different applications. To insist on taking full ownership and responsibility for it, instead of letting third-party vendors control it. To commit to investing time, money and resources into realizing its full potential. The Marketing Data Lake is an enormously powerful tool around which to build a new kind of marketing operation that rises to the challenges and opportunities represented by your data. But it's only a tool. As a marketer, it's your job to harness it and to build the processes around it that squeeze out the value. To find data-oriented people to work the machine. And stakeholders to harvest the rewards. And executive sponsors to pave the way and watch your back. This may sound like an improbable, frightening road. But we've traveled it and we can report that it's amazingly rewarding. More importantly, the alternative – blind, opaque marketing fragmented into dozens of silos – is simply unsustainable. We hope this book has helped you think about your discipline, your job and your career differently. And we hope it's inspired you to make the leap to completely data-driven marketing. Thank you for reading. Go forth and geek out.

Appendix: Who should NOT build a Data Lake

We don't want to pretend that a Marketing Data Lake is the answer for everyone all the time. If you fall into any of the categories listed here, you may not be right for (or ready for) a Data Lake approach:

People who need penny-perfect data
If you need 100 percent perfection – for financials or compliance or research – the Data Lake may not be ready for you. One day, it will be, but today it may not be mature enough for the finest grain detail.

People who only work with structured data
If all you'll ever work with is highly structured data – from transactions, for instance – then a traditional data warehouse may be just fine for you. You may not get full value from the Data Lake approach (unless you like the idea of adding in new sources…).

People with zero IT support
If you're not particularly techie and you don't have the support of your IT team, don't try this at home. You will need IT people to get

yourself set up (after that, you'll need them far less). If your IT folks are too busy or they hate the idea… pause for a re-think.

People who aren't into analytics
If your data is for in-line transaction processing and things like that, you don't need a Data Lake. If you never expect to analyze the data, put your resources elsewhere.

People in most small companies
The Data Lake is ideal for larger enterprises with lots of complexity, many data sources and probably a fair amount of data volumes. Most small companies can do a lot simply by optimizing within their existing systems, extracting data for offline analysis. Some small companies – the very data intensive ones – may still benefit from their own Lake.

People whose core marketing systems are a mess
If your web analytics, CRM or marketing automation systems are in chaos, you need to fix that first before even thinking about combining their sludge into a central Lake. Do that, then you're good to go.

People who are ready to retire
Hey, if you're at the end of your career, you probably don't need to climb up this particular learning curve. Get the whippersnappers to do it. (Unless you're the insatiably curious type.)

Not on this list?
Welcome aboard!

Thank yous

Thanks to our amazing Informatica colleagues for their help, patience, encouragement, enthusiasm and intelligence. Without our friends in Sales, IT, Product, Marketing and the C-Suite, we'd never have reached first base.

Thanks to the entire Informatica marketing team for putting so much into this transformation – and doing it with such good humor and grace. You're the best.

Thanks to our families for listening to our insane babble night after night, weekend after weekend. We love you!

Thanks to the excellent teams in our supplier companies and agencies – always ready to hop on a call or pop into the office. You're awesome.

Thanks to Scott Brinker, Adam Greco and Doug Kessler – for your contribution to this book and to the discipline.

Finally, thanks to the front-line practitioners – the Informatica customers and users who share their experiences generously with us on user forums, at Informatica World events and in person every day. We promise to work our butts off to reward your confidence in us!

Further reading

Books

- HACKING MARKETING: AGILE PRACTICES TO MAKE MARKETING SMARTER, FASTER, AND MORE INNOVATIVE, by Scott Brinker
- THE ADOBE SITE CATALYST HANDBOOK: AN INSIDER'S GUIDE, by Adam Greco
- BIG DATA: USING SMART BIG DATA, ANALYTICS AND METRICS TO MAKE BETTER DECISIONS AND IMPROVE PERFORMANCE, by Bernard Marr
- STORYTELLING WITH DATA: A DATA VISUALIZATION GUIDE FOR BUSINESS PROFESSIONALS, by Cole Nussbaumer Knaflic
- DATA INTEGRATION FOR DUMMIES, available on the Informatica website
- REVENUE DISRUPTION: GAME-CHANGING SALES AND MARKETING STRATEGIES TO ACCELERATE GROWTH, by Phil Fernandez

Websites & Resources

- **Chiefmartec** insights on marketing technology
- **Marketing Tech Blog** insights from Douglas Karr and colleagues
- **MarTech Advisor,** research, news, reviews
- **Analytics Demystified,** on all things analytics
- **The Content Marketing Institute** for guidance on content marketing
- **Moz.com** for guidance on SEO
- **Sirius Decisions,** the research and advisory firm

First published in the United States of America 2016.

ISBN 978-0-9976012-0-6

verbal books
London and New York

Layout & design: Melanie Bartheidel