BIG DATA IN PRACTICE

"A comprehensive compendium of why, how, and to what effects Big Data analytics are used in today's world"
James Kobielus, Big Data Evangelist at IBM

FROM BIG DATA TO REAL BUSINESS VALUE:
THE SMART APPROACH AND 3 MORE USE-CASES

BERNARD MARR

WILEY

FROM BIG DATA TO REAL BUSINESS VALUE:

THE SMART APPROACH AND 3 MORE USE-CASES

CONTENTS

SMART Data vs. Big Data 5
The SMART model at work 6
SMART questions, smart answers 7
A collaborative exercise 8
Linking Big Data with strategic goals 10

Case Study No. 1:
How Big Data Means Hot Pizzas for Domino's 12

Case Study No. 2:
The NFL Understands How to Play the Big Data Game 15

Case Study No. 3:
NASA 19

Everyone in business knows about big data now, and how it's going to revolutionize the world we live in. Information is a valuable commodity – the more we have of it, the more we can learn from it, and make changes that will drive business success.

That's the theory anyway – and what many people will tell you if you ask them to define "Big Data".

But in the rush to avoid being left behind, I also see that many companies, frantically measuring and recording everything they possibly can, risk becoming "data rich but insight poor". In short, too many have gathered a lot of data they have no idea what to do with, and no hope of learning anything useful from.

"BUSINESSES USE LESS THAN 5% OF THE INFORMATION AVAILABLE TO THEM TO INFORM DECISIONS"

To add to the problem, a lot of data has a lifespan. At some point in time, it becomes no longer relevant – inaccurate or outdated. But often it is held onto it anyway – with vague thoughts that it may one day prove useful in some way. Storing all this data has its cost. Ignoring the costs of collection, data requires somewhere to keep it, electricity to power it and, if your information is sensitive (such as customer records) time and money to be spent on security and data compliance.

In fact, research has shown that on average, businesses use less than 5% of the information available to them to inform decisions.

To be fair to them, it isn't entirely their own fault – you see, the problem is that a lot of businesses have been misled. And much of the blame for that lies with Big Data itself. Specifically, its name.

In fact, Big Data isn't about size at all. As the old saying goes, it's not how big it is, it's what you do with it that counts. In my opinion, we shouldn't be thinking about Big Data at all. That may sound like a strange comment coming from someone who works in Big Data for a living. But, I believe what we should be focusing on, is Smart Data.

Those who have really profited from the explosion in available data, and our increasing expertise at analyzing and learning from it are those who have learned to "think smarter" about data and what to do with it. Industry-dominating behemoths such as Google, Facebook and Amazon have all been pioneers – not simply collecting vast quantities of data, but finding pioneering ways to put it to use.

Sure, these are companies with vast resources at their disposal – millions of users willing to share increasingly intimate details of their lives, as well as teams of highly-trained data scientists versed in the arcane arts of turning data into insights. But one thing they have in common – something by no means only available to multinational corporations which turn over billions – is they have been smart about how they have used their data.

And that forms the foundation of my SMART Data framework – which you can use to move from thinking about Big Data and instead think about smart data.

Once you understand the difference between just Big Data and smart, actionable data, you'll see how three successful — and wildly different — companies are putting data to use in their organizations. Domino's Pizza, the NFL, and NASA are all making innovative use of the data at their disposal to move their business forward.

SMART Data vs. Big Data

We have been told for years that we live in the Information Age: Information is gathering momentum and pace; it's growing exponentially.

The escalation of data poses its own set of problems — the danger is we get lost in a sea of data that delivers no value whatsoever, bamboozled by the proliferation of smart technology and endless possibilities that send businesses down resource-sapping rabbit holes without any useful or definable output.

The basic idea behind Big Data is that everything we do is increasingly leaving a digital trace that we can use and analyze to become smarter. US retailing giant Wal-Mart Stores Inc. handles more than 1 million customer transactions an hour and imports them into databases estimated to contain more than 2.5 petabytes of data.

Wal-Mart uses data to understand what is trending in social media, as well as buying patterns amongst similar types of customers and what competitors are charging in real time. The company changed its free shipping policy based on Big Data analytics, increasing the minimum qualifying online order from $45 to $50. Meanwhile, Amazon has recently patented "anticipatory shipping". The online retailer has become so good at predictive analytics that it believes it knows what you will buy before you buy it and can ship an item towards you even before it is in your shopping cart.

Companies such as Amazon and Wal-Mart may enjoy a competitive advantage because of the data they have access to, but there is probably more than enough data in your own company to tap into without eye-watering budgets as long as you focus on SMART data, not Big Data.

The SMART model — start with strategy, measure metrics and data, apply analytics, report results, transform your business — allows you to cut through the chaos. Step back from the hype and learn the techniques to convert raw information into real business results.

The SMART model at work

Instead of starting with the data, start with your business objectives and what you are specifically trying to achieve. This will automatically point you towards questions that you need to answer, which will narrow data requirements into manageable areas.

Once you know what you are trying to achieve and you are clear on what SMART questions need to be answered, then work out how you can access that information so you can measure metrics and data.

The next step is to apply analytics, extracting useful insights from the data that can help you answer strategic questions. The data themselves are meaningless unless they help you to execute your strategy and improve performance.

But the insights alone are useless unless you report results. The mistake that is too often made by data scientists is to bury these insights in 50-page reports that no manager or executive has time to read. Ideally, results should also be reported in a single-page infographic.

When you approach data and analytics from this narrower, more practical perspective, you can transform your business, understanding your customers better, improving and optimizing business processes, and driving performance.

I worked with a telecom company whose business objective was to reduce customer churn. Executives knew they had a lot of data and ran analytics on some of it but had never looked at how people called each other. They didn't know whether their customers made mainly inbound or outbound calls, how long they spoke, or what times of the day were most popular.

By mining those data and applying analytics, they found that one particular calling pattern was much more associated with churn than the rest. The data analysis showed that there was a type of customer who moves more frequently than others. By identifying that segment of the market, they could target those customers with special offers that would entice them to stay. By analyzing traditional structured data a little differently, they were able to extract commercially significant insights that increased profit.

SMART data can transform the relationship of a business with its customers. They can also transform the way a business thinks about itself.

SMART questions, smart answers

Asking SMART questions can lead a company to change its business model entirely.

Rolls Royce asked itself what its customers needed rather than focusing simply on what the company provided. That led it from only manufacturing aircraft engines to monitoring the engines, too, using thousands of sensors to identify issues before they arose.

Now Rolls Royce sells the engines and offers to monitor them, charging customers based on engine usage time, and repairs and

replaces parts if there is a problem. So the client effectively buys a dynamic servicing option, and this servicing now accounts for 70% of the civil-aircraft engine division's annual revenue.

General Electric Co. is another company that has transformed its business away from its traditional manufacturing roots using Big Data analytics. Its gas turbines are now smart turbines. It has collected 100 million hours of operational data, and these "intelligent" machines communicate their operating statistics through 400 sensors on each turbine, allowing adjustments to be made to improve efficiency and reduce wear and tear.

Using operational data from sensors on a range of machinery and engines, GE applies analytics to identify patterns and deliver commercially relevant insights. Like Rolls Royce, GE provides additional services tied to its products designed to improve real-time efficiency and minimize downtime caused by parts failures.

As well as transforming a company's relationship with its customers, Big Data is increasingly being used to optimize business processes.

A collaborative exercise

To reap the benefits of Big Data, you don't have to collect everything and produce the biggest, most complex database in the world. The aim is actually the opposite — to get clear about the data you need and build the smallest, most straightforward database in the world.

The truth is we are often so mesmerized by the data that we forget the question is much more important than the answer the data may provide.

As a rule of thumb, start with internal data and structured data, which are usually easier and cheaper to analyze than unstructured or semi-structured data.

But asking the right questions and finding the right data to answer them is not in itself enough. Leaving the reporting to analysts and designers alone is as unproductive as leaving it to the executives. There needs to be collaboration and interaction between the people creating the results and the people who need the results to make decisions.

Consumer products manufacturer Procter & Gamble has chosen to institutionalize data visualization as a management tool. The company put visual displays of key information on more than 50,000 desktops and created "business spheres", which are basically information dashboards that have all the same technology and data visualization protocols in place.

P&G has ensured information is presented in a common way across the whole company. The company has initiated a set of seven "business sufficiency models" that specify what information is used to address particular problem domains. If a P&G executive is focused on supply-chain issues, the sufficiency models specify the key variables, how they should be displayed visually, and sometimes even the relationships between the variables and forecasts based on the relationships.

The uniformity means everyone is on the same page. Once the data are understood, executives from any division, any brand, or in any country can quickly and easily interpret the data. That means they spend less time trying to understand the data and more time putting them to use and making better decisions. It also prevents people

from hijacking the data and presenting them in a way that supports their pet theory.

Never forget that data analytics is useful only if you make sure the right people get the right information in the right format so they can make the right decisions more often.

Linking Big Data with strategic goals

To help companies get clear on their strategic data needs, I've developed a framework I call the SMART strategy board. It helps you step back and ask the right questions to ensure your Big Data initiatives are linked to your strategic needs.

There are six panels in the strategy board (see Figure 1):

1. **The Purpose Panel** contains your mission statement, which concisely states why your business exists, and your vision statement, which explains what you want your business to be in the future.
2. **The Customer Panel** prompts you to consider how much you know about the customers your strategy is targeting and what you may need to find out in order to deliver on your strategic objective. It should include your target market and your value proposition.
3. **The Finance Panel** is designed to make you consider how much you know about the financial implications of your strategy. How does your strategy generate money? Are you confident your business model is accurate?
4. **The Operations Panel** makes you think about what you need to do internally to deliver your strategy. Which suppliers, distributors, and partners are crucial? Do you currently work with them, or do you have to establish new relationships?

And, secondly, in what core competencies will you need to excel if you are going to execute your chosen strategy?

5. **The Resource Panel** prompts you to consider what resources you need to deliver your strategy and what you may need to find out. There are four components: IT systems and data; infrastructure; people and talent; and values and leadership.

6. **The Competition and Risk Panel** assesses the competition you will be up against and what risks you may face along the way. It's the perspective that is most often missing from strategy maps, and yet it poses a serious potential threat to successful strategic execution.

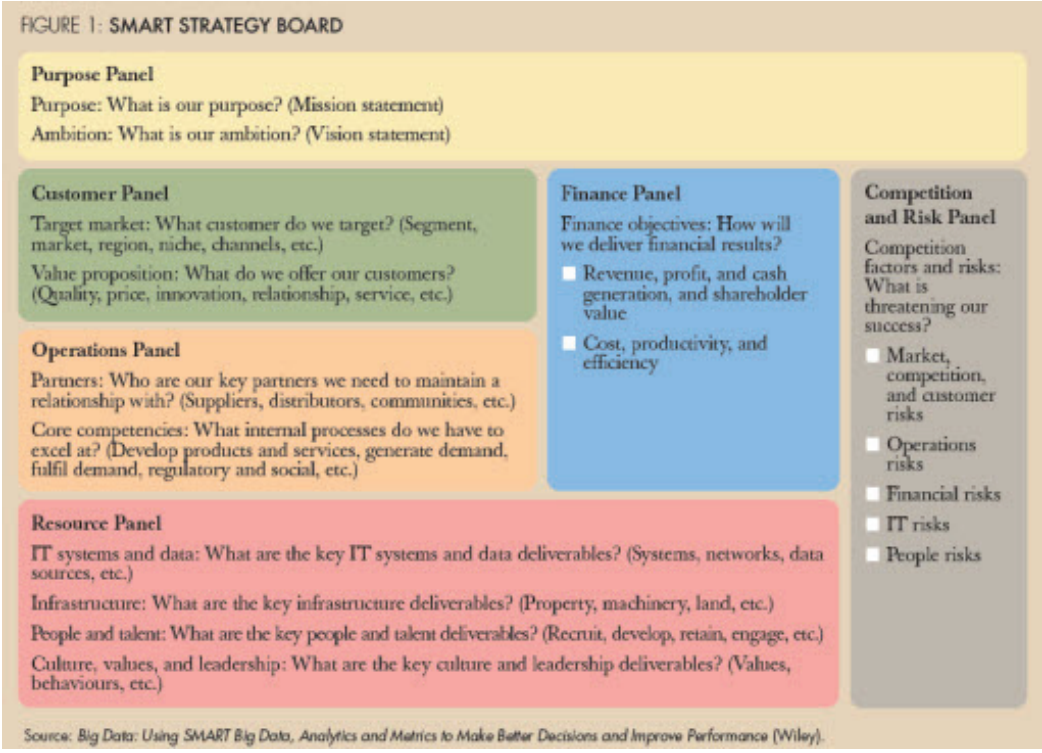


Figure 1.

CASE STUDY NO. 1:

How Big Data Means Hot Pizzas for Domino's

Pizza delivery might not seem, on first glance, the most high-tech of industries. But Domino's, the world's largest delivery chain, has consistently pushed its brand towards new and developing tech, and as such it is now possible to order pizzas on Twitter, smartwatches and TVs, in-car entertainment systems such as Ford's Synch, and social media platforms like Facebook.

The initiative to keep a Domino's order button at the fingertips of their customers at all times is called Domino's AnyWare. Of course, this multi-channel approach to interfacing with customers gives rise to the potential to generate and capture a lot of data – which Domino's capitalizes on by using it to improve the efficiency of their marketing.

I had the chance to talk to data management lead Cliff Miller and Dan Djuric, vice president of enterprise information services, about how the business uses Big Data to derive insights on customers, as well as how it sees analytics as the key to driving change across the industry.

"Domino's AnyWare literally translates to data everywhere," Djuric tells me.

Data captured through all its channels – text message, Twitter, Pebble, Android, and Amazon Echo to name just a fraction – is fed into the Domino's Information Management Framework. There it's combined with enrichment data from a large number of third party sources such as the United States Postal Service as well as geocode information, demographic, and competitor data to allow in depth customer segmentation.

"Ultimately we start to build this unified customer view, measuring consistent information across our operational and analytic layers," Djuric says. "So this is now where we start to tell the story."

Information collected through the group's point of sales systems and enrichment data add up to 85,000 data sources, both structured and unstructured, pouring into the system every day.

The family or household is fundamental to Domino's tactics for segmenting customers. "Pizza ordering is a household exercise," says Djuric.

"We have the ability to not only look at a consumer as an individual and assess their buying patterns, but also look at the multiple consumers residing within a household, understand who is the dominant buyer, who reacts to our coupons, and, foremost, understand how they react to the channel that they're coming to us on."

This means that individual customers or households can be presented with totally different presentation layers than others – different coupons and product offers – based on statistical modelling of customers fitting their profile.

Customer segmentation data is also used to assess performance and drive growth at individual stores and franchise groups. "We can work with a store to tailor their coupons, and tailor what we think their upselling capabilities might be in terms of adding more revenue.

"But what we can also do is tell them what is not working for them – considering not only their market but comparing their market to other markets, and also to our competitive landscape."

Domino's investment in driving orders across their digital platforms has certainly borne results – the group now processes about 55% of its orders via online systems, with the remainder accounted for by traditional ordering methods such as telephone or in-person.

"We're becoming more of a digital e-commerce operation, rather than a traditional quick service restaurant business," Djuric tells me, "We're

leading the way, I think, in what we're doing with social and what we're doing with our digital platforms. And importantly, I think, it's structuring the digital service with the insights we have on customers and practices, that leading our customers to a better, faster and more quality-based experience on our digital platforms. It has surprised us, how quickly we've transitioned from a traditional ordering mechanism into an e-commerce based profile."

Domino's data management lead Cliff Miller was also keen to talk about how influential analytics and Big Data has been in that transition.

"We want to be sure we're getting the best product we can to our customers as fast as we can, but we always want to analyze all segments of our business, as we believe being a data driven company gives us a lot of competitive advantages."

"We've definitely moved away from a time when we were just processing data on sales and operational metrics in our warehouse, every business unit under our umbrella is looking to leverage data to make them faster and more cost effective."

Domino's relies on Talend's Big Data Integration platform for much of the heavy back-end work. across Enterprise Data Warehouse .

"We've already got a ton of stuff in Talend, and it's become our de-facto enterprise data processing platform, it's extremely flexible in terms of what it can do."

"We're an aggressively aspirational company and store growth is very important. We need to make sure we are on platforms which can scale."

Domino's is clearly a business which is not scared to move with the times. From pioneering online and TV-based ordering to advanced customer analytics, it has consistently moved to keep with the latest trends in technology. The logistics of delivering close to a million pizzas a day across 70 countries throws up exactly the sort of problems

that Big Data is good at overcoming. The company has recently been experimenting with new delivery methods – such as a car with a built in oven capable of delivering 80 pizzas in one journey, in order to cut the carbon footprint of deliveries. It will be interesting to see where their enthusiasm for technology and analytics takes them in the future.

CASE STUDY NO. 2:

The NFL Understands How to Play the Big Data Game

Like many businesses, the National Football League is experimenting with big data to help players, fans, and teams alike.

The NFL recently announced a deal with tech firm Zebra to install RFID data sensors in players' shoulder pads and in all of the NFL's arenas. The chips collect detailed location data on each player, and from that data, things like player acceleration and speed can be analyzed.

The NFL plans to make the data available to fans and teams, though not during game play. The thought is that statistics-mad fans will jump at the chance to consume more data about their favorite players and teams.

In the future, the data collection might be expanded. In last year's Pro Bowl, sensors were installed in the footballs to show exactly how far they were thrown.

Of course, this isn't the NFL's first foray into big data. In fact, like other statistics-dependent sports leagues, the NFL was crunching big data before the term even existed.

However, in the last few years, the business has embraced the

technology side, hiring its first chief information officer, and developing its own platform available to all 32 teams. Individual teams can create their own applications to mine the data to improve scouting, education, and preparation for meeting an opposing team.

It's also hoped that the data will help coaches make better decisions. They can review real statistics about an opposing team's plays or how often one of their own plays worked rather than relying solely on instinct. They will also, in the future, be able to use the data on an individual player to determine if he is improving.

Diehard fans can, for a fee, access this same database to build their perfect fantasy football team. Because, at heart, the NFL believes that the best fans are engaged fans. They want to encourage the kind of obsessive statistics-keeping that many sport fans are known for.

It's hard to predict how this flood of new data will impact the game. In 2015, only 14 stadiums and a few teams were outfitted with the sensors. And in 2016, the NFL decided against installing sensors in all footballs after the politics of 2015's "deflate gate" when the Patriots were accused of under inflating footballs for an advantage.

Still, it seems fairly easy to predict that the new data will quickly make its way into TV broadcast booths and instant replays. Broadcasters love to have additional data points to examine between plays and between games.

2016 also marked the 50th year of the Super Bowl, and plenty has changed since the first time teams duked it out for the national title. Sports analysts have been collecting data on football games since the very beginning, but our data collection techniques and abilities have vastly improved. New sensors in stadiums and on NFL players' pads and helmets help collect real-time position data, show where and how far players have moved, and can even help indicate when a player may have suffered a damaging hit to the head.

In 2013, Microsoft struck a $400 million deal with the NFL to make this data available to coaches and players via their Surface tablets. Coaches use the tablets to demonstrate and review plays on the sidelines, as well as access real-time data from the NFL's databanks.

Of course, the partnership hasn't been all smooth. At first, commenters kept referring to the tablets as "iPads," and had to be reminded many times to call them by their brand name. Then, during the AFC championship game this year, the Patriots briefly had connectivity issues, causing the tablets to stop working.

And, of course, the commenters remembered to call them by their brand name that time just as, almost as unfortunately, the network ran a Microsoft ad during the break.

Still, connectivity issues aside, the advent of real-time data access is potentially changing the way coaches view games and call plays.

As data collection has improved, people are also turning more and more to computer algorithms to predict the outcome of the game.

Big data, IoT sensors, and new analytics abilities makes even more data available, including real-time distance traveled and position on the field, how weather conditions affect individual plays, and even predicting individual player matchups.

But it's still a long way from a sure bet.

A company called Varick Media Management has created its own Prediction Machine to predict the outcome of all the games in the season. The site also offers its "Trend Machine" which can analyze many different matchups from more than 30,000 games over 35 years of play.

However, their accuracy is far from perfect; in the 2014-2014 regular

season, they boasted a 69% accuracy rating. But, they were the only major predictor to foresee the Seattle Seahawks blowout win over the Denver Broncos in 2014. Facebook tried to use social data to predict the outcome, and predicted the Broncos would take home the trophy.

Another way big data could be changing the game is in the realm of advertising.

Everyone knows Super Bowl ads cost millions of dollars — a record $5 million for a 30-second spot this year — but what big data showed marketers last year is that most of the online chatter about Super Bowl XLIX took place after the game.

This could be a boon for advertisers who want to take advantage of the attention on the game, but can't afford a TV spot. In fact, by using data to strategically target only certain consumers, online advertising could represent a much better investment with better engagement — and at a fraction of the cost.

That could mean that, for well targeted ads, online advertising could have a better return on investment than television, even ads aired during the most-watched event on live TV.

In addition, social listening in years past told us that **social media users talked more about the brands**, commercials, and halftime show than they did the game. That points to an important social opportunity for brands who can afford to advertise during the game.

As we all know, brands can make or break a big social opportunity like this depending on how well their social media is managed, and how well they're watching their data. Take the **Oreo blackout tweet during the 2013 Super Bowl** — many said the Twitter ad "won" the Super Bowl marketing game that year, and they paid mere pennies compared to brands that bought TV ads.

CASE STUDY NO. 3:

NASA's Mars Rover – Real Time Analytics 150 Million Miles From Earth

The team in control of NASA's Mars Rover spacecraft now have Big Data driven analytical engines at their fingertips.

The same open source ElasticSearch technology used by the likes of Netflix and Goldman Sachs is now being put to use planning the actions of the Curiosity rover, which landed on the red planet in 2012. And the next mission, due to launch in 2020 with the primary mission of finding evidence of ancient life on the planet, is being built around the technology from the ground up.

NASA's Jet Propulsion Labs, which runs the day-to-day mission planning, has rebuilt its analytics systems around ElasticSearch which now processes all of the data transmitted from the Rover during its four daily scheduled uploads.

These uploads cover tens of thousands of data points – every reading taken by Curiosity's onboard sensors including temperature on the Martian surface, atmospheric composition, and precise data on the rover's equipment, tools and actions.

All of Curiosity's operations are planned a day in advance based on data received the previous day. This move to a real time analytics vastly speeds up the time in which decisions can be taken by mission control.

ElasticSearch – which recently achieved its 50 millionth download – means that patterns and anomalies in the datasets can be spotted far more quickly. Correlations which could provide mission-critical insights are more likely to become apparent, leading to a greater rate of scientific discovery and less danger or malfunction or failure.

NASA JPL data scientist Dan Isla told me "This has been very transformational from an operational point of view. We can get very quick insights over multiple days' worth of data, or even the entire mission, in just a few seconds, without delays. We can actually ask questions and get answers quicker than we can think up new questions."

One application is anomaly resolutions. When it appears that there is a problem with the spacecraft, precise details of its operations can immediately be analyzed to find out when the last time this situation occurred, and what other elements were in play at the time.

The system is also used for the vital function of managing power generated by the rover's radioisotope thermoelectric and solar power generators. The trickle charge system only makes around 100 watts of power available to the craft and its instruments at any one time consistently.

This means that power management must be carefully planned in advance. As Isla explains, "using ElasticSearch we can budget this power much more effectively, we can look at the data and see 'ok, this is what we used yesterday, and this is the state of the batteries, and this is our plan coming up…' we can build a power model very quickly.

"With ElasticSearch everyone can do analytics, ask questions at the same time and get the same results, we're all working together now. Everyone's very agile now."

This isn't NASA's only project to use ElasticSearch. It was incorporated into the Mars Rover program after being successfully tested in the Soil Moisture Active Packing mission, launched last year. SMAP uses high powered radar and radiometer to analyze and capture high resolution data on soil moisture across the globe. A full moisture map of the Earth's surface is generated every three days – data creation an order of magnitude greater than that of the Mars Rover. While the rover has

so far generated around 1 billion data points, the SMAP project is already pushing 20 billion.

Perhaps the most exciting aspect of Mars exploration today is the possibility of discovering whether or not the planet was ever home to life. It is thought that ElasticSearch will greatly speed up the answering of this question.

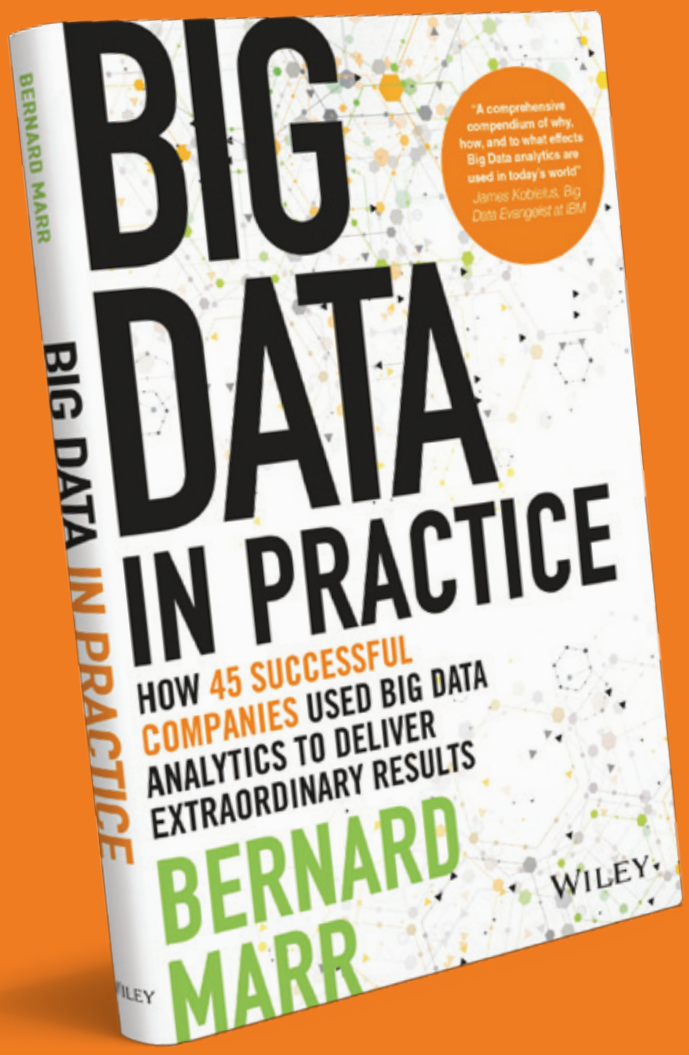
Isla tells me "Curiosity was designed to find out whether Mars was ever a habitable environment. During the course of the mission we've been able to confirm that yes, it was. The next mission – in 2020 – we're taking life detection instruments and we need more capabilities, more instruments and we need to be able to do it faster.

"We can't be spending hours analyzing the data, so we need to take the ElasticSearch approach that we implemented fairly early into Curiosity and do that from day one. We're definitely baselining this tech as our operational data system."

WHAT'S NEXT:

If you enjoyed the insights from these fascinating case studies, you're going to love the 45 additional case studies I present in ***Big Data in Practice***.

If you want more information about the SMART formula for Big Data, I explain it in much more detail in my previous book, ***Big Data: Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance***.



amazon.co.uk amazon.com

ABOUT THE AUTHOR

Bernard Marr is a leading global authority and best-selling author on organizational performance and business success. He is a LinkedIn Influencer, he writes the Big Data Guru blog and is the world's #1 expert on big data. He regularly advises leading companies, organizations and governments across the globe, and is an acclaimed and award-winning keynote speaker, researcher, consultant and teacher.

Bernard is the founder and CEO of the Advanced Performance Institute. Prior to this he held influential positions at the University of Cambridge and at Cranfield School of Management. Bernard Marr's expert comments on organizational performance have been published in *The Times*, *The Financial Times*, *The Sunday Times*, *Financial Management*, the *CFO Magazine* and *The Wall Street Journal*.

Connect with Bernard