Making Everything Easier!™

Big Data For Small Business

FOR DUMMIES®

A Wiley Brand

Learn to:

- Use big data to make better decisions
- Transform your business model
- Understand the technology
- How to turn data into insights

Bernard Marr
Founder and CEO, Advanced Performance Institute

Introduction

Almost everything we do now leaves a digital trace. If you bought this book online, you left a trail of digital crumbs in your wake, from browsing the online retailer's website, to the transaction itself. Even if you walked into a physical bookstore and paid with cash, there's still likely to be a digital trail of your activities, including CCTV (closed-circuit television) footage and location data from your own phone.

These digital traces can be summed up in one phrase: big data. *Big data* refers to the ability to collect and analyse the vast amounts of data now being generated in the world. This ability to harness the ever-expanding amounts of data is completely transforming our ability to understand the world and everything within it – from healthcare and science to how entire cities and countries are run. And, of course, it's transforming the way we do business.

Some business owners and managers dismiss big data as being only for big-budget corporations. I think this is a huge mistake. Of course it's true that some companies have eye-watering budgets for big data analytics, but most simply don't. In fact, I work with plenty of small- and medium-sized businesses that successfully harness the power of data without spending a fortune.

The key is to start with a clear strategy. This allows you to focus solely on the data that's right for you – the data that will help you achieve your long-term business goals. Having a clear strategy helps you cut through the hype and noise surrounding big data and get straight to how it can realistically help you improve the way you do business. That's why I wrote *Big Data for Small Business*

For Dummies: to help SMEs (small and medium enterprises) use big data in a practical and strategic way.

Whether you're planning a one-off data project or want to incorporate data into your ongoing business operations, this book can help you understand what big data is, how you can apply it to your business, how to create your own big data strategy and get underway and how to build a culture that emphasises data-based decision making and continuous improvement.

About This Book

Think of this book as a no-nonsense tour guide to help you on your big data journey. There are lots of inspirational examples of how other businesses are already using data, but the focus remains on practical tips to get you using data in *your* business. As well as examples and tips, the book is packed with step-by-step guidelines and lists designed to help you get the most out of big data. All the information is designed to be accessible and easy to understand. And where I have to resort to technical jargon, I give clear definitions. Sidebars (the grey boxes) contain nice-to-know but not essential information, so you can easily skip over them if you like.

The book is designed as a resource that you can dip in and out of and return to time and time again. As such, you don't need to read it from cover to cover (although, if you want to, go ahead!). It's designed to be read in whatever order works best for you.

Finally, if you decide to visit a website listed in the book then you just need to copy the URL (uniform resource locator) exactly as it appears in the book. This is true

even if the address falls between two lines or two pages – no extra characters (such as hyphens) were inserted.

Foolish Assumptions

Every author has a target audience in mind when he writes. For this book, I assume that you're the owner of a small/medium business or a manager in such a business. I assume that you've heard a little about big data already – perhaps what a powerful tool it can be for businesses – and you want to know more. I don't assume you have any prior technical knowledge whatsoever. Crucially, whether you're a business owner or manager, I assume that you want to improve the way you do business and you're in a position to make strategic decisions … and then act upon them.

If you would like to supplement this book with more technical information, you might like to check out *Big Data For Dummies* by Judith Hurwitz, Alan Nugent, Fern Halper and Marcia Kaufman, published by Wiley.

Icons Used in This Book

When I see a huge wall of text, I start to switch off. So in this book I use a number of icons to break the text up, and to make it easier and more enjoyable to read. The icons also help you spot key information quickly.

These indicate expert advice or suggestions to help your big data journey go more smoothly. They help you save time, energy or money and are based on my experiences working with other businesses.

This icon flags critical material that you should store away in your memory for later use. But don't worry – they're usually very short.

As the name suggests, this icon flags potential pitfalls that you need to avoid as you start using big data.

Where I've had to use data-related jargon (for example, if it's a key industry term that you need to know), I highlight it using this icon.

I believe real-life examples bring information to life and I included tons of examples throughout the book that show how other businesses are already using big data. Some of these are small-scale data projects, others are much larger, and some are just plain weird and wonderful! But all are designed to demonstrate the exciting potential of big data and give you a few practical ideas for your own business.

Beyond the Book

In addition to the material in the print or e-book you're reading right now, this product also comes with some access-anywhere goodies on the Web.

Check out the free Cheat Sheet at www.dummies.com/cheatsheet/bigdataforsmallbusiness for some helpful key information and checklists. It's designed as a

quick-check reference for some crucial big data information, including a handy list of key terminology.

There are also some useful bonus articles and an additional Part of Tens chapter available on the website. Head to www.dummies.com/extras/bigdataforsmallbusiness to access these.

You may also like to check out the website of the Advanced Performance Institute, which I founded and head up. There you'll find many relevant case studies, white papers and reading material on big data: www.ap-institute.com. I also write regularly for *Forbes* magazine on all things big data and you can find my articles at www.forbes.com/sites/bernardmarr. My LinkedIn page also contains a wealth of articles and posts on big data: www.linkedin.com/in/bernardmarr.

Where to Go from Here

The short answer is: It's up to you. You don't have to start at Chapter 1 and work your way through the book in a linear way – but you can if you want to.

If you're completely new to big data, I recommend you start with the Chapters in Part I for an explanation of what big data is and the main ways it can be used in business. Otherwise, simply use the table of contents to find what you're most interested in and jump straight to that section. If you want to start by finding out exactly how to create a big data strategy, turn to Chapter 10. If you're interested more in big data skills and competencies, start with Chapter 8. Wherever you go from here, you'll find a wealth of information and tips to help you start using big data in your business.

Part I
Getting Started with Big Data Basics

For Dummies can help you get started with lots of subjects. Go to www.dummies.com to find out more and do more

with *For Dummies*.

In this part …

- Understand what big data is and why you need to know about it.
- Find out the key characteristics that define big data.
- See why there’s so much hype around big data right now – and why all the fuss is justified.
- Check out key big data uses for small businesses.

Chapter 1

Introducing Big Data for Small Businesses

In This Chapter

- Understanding what big data is and why it's so important
- Taking a peek at the different types of data
- Putting big data to work in your business

Big data has been making big headlines over the last couple of years, but it's much more than just a buzz phrase or the latest business fad. The phenomenon is very real and is producing concrete benefits in so many different areas – from business to medical research to national security.

The basic idea behind big data is that everything you do is increasingly leaving a digital trace (or data), which you (and others) can use and analyse. *Big data* therefore refers to that data being collected and the ability to make use of it.

In this chapter, I look at how this phenomenon is transforming the way you do business. I also look at what sorts of data are available these days and introduce my step-by-step processes for using big data in business.

Personally, I don't love the term *big data* because I think it places far too much emphasis on the sheer volume of

data, when, as I talk about in this chapter, what you do with the data is much more important than how much of it you have. I have a feeling the term will gradually disappear and what's now called *big data* will, in the future, just be known as *data.*

Why Big Data Matters to Every Business, Big and Small

Given all the hype around big data, it's no surprise that market researchers Gartner found in 2014 that 73 per cent of businesses have already invested in a big data plan or are planning to do so in the next few years.

The online behemoths that have come to dominate business in the Internet age – Google, Facebook, Amazon, you know the ones – all base their business models on big data. It's by collecting and analysing huge amounts of information from us that they're able to determine precisely what we want. The data also enables them to sell advertising services capable of precisely targeting their clients' preferred demographics.

But big data isn't just for giant corporations, it matters to every company – no matter how small or traditional. To cater for this huge demand, many companies have sprung up to offer services to other businesses, enabling them to launch big data initiatives of their own.

I've found that the term *big data* can scare some business owners off, especially the *big* part. Some think it doesn't apply to their small business – or that

it'll be prohibitively expensive to use. The truth is, in the current business age, if you want to grow, you need to start using data strategically. And yes, some companies have massive big data budgets. Most companies, however, are working with much smaller budgets but are still able to use data to gather insights that help their business grow.

Entering a world filled with data

Of course, data collection itself isn't new. But technological advances like chip and sensor technology, the Internet, cloud computing and the ability to store and analyse data have changed the quantity of data you can collect.

Activities that have been a part of everyday life for decades – shopping, listening to music, taking pictures, talking on the phone – increasingly happen wholly or in part in the digital realm and therefore leave a trail of data.

And the amount of data being generated every day is staggering. For example, users of Facebook upload around one billion pieces of content to the social network site every day. In industry, machinery and vehicles are fitted with sensors and trackers that record their every move, and whenever you call a call centre, an audio recording of your conversation is made and stored in a huge digital database.

Big data statistics to blow your mind

The following statistics give you a flavour of the sheer volume of data being generated in today's world:

- Every two days we create as much information as we did from the beginning of time until 2003.
- Over 90 per cent of all the data in the world was created in the past two years.
- It's expected that by 2020 the amount of digital information in existence will grow from 3.2 zettabytes today to 40 zettabytes. (A zettabyte is a unit of data measurement roughly equal to one million terabytes; a terabyte is one trillion bytes.)
- Every minute we send 204 million emails, generate 1.8 million Facebook likes, send 278 thousand Tweets and upload 200 thousand photos to Facebook.
- Google alone processes on average over 40 thousand search queries per second, making it over 3.5 billion in a single day.
- Around 100 hours of video are uploaded to YouTube every minute and it would take you around 15 years to watch every video uploaded by users in one day.
- 570 new websites spring into existence every minute of every day.
- This year, there will be over 1.2 billion smartphones in the world (which are stuffed full of sensors and data collection features), and the growth is predicted to continue.
- The boom of the Internet of Things will mean that the amount of devices that connect to the Internet will rise from about 13 billion today to 50 billion by 2020.
- 1.9 million IT jobs will be created in the US alone by 2015 to carry out big data projects. Each of those will be supported by three new jobs created outside of IT – meaning a total of 6 million new jobs thanks to big data.

Understanding the infinite ways to use big data

Eventually, every aspect of your lives will be affected by big data. However, there are some areas where big data is already making a real difference today – in business and in other areas. Let's look at the main areas where big data is most widely used right now.

- **Understanding and targeting customers:** This is one of the most common uses of big data today. Here,

big data is used to better understand customers and their behaviours and preferences.

- **Understanding and optimising business processes:** Big data is also increasingly used to optimise a wider range of business processes, including stock control, supply chain and delivery routes and HR processes.
- **Optimising personal quantification and performance:** Individuals can now benefit from the data generated from wearable devices such as smart watches and smart bracelets – data like calorie consumption, activity levels and sleep patterns.
- **Improving healthcare and public health:** The computing power of big data analytics enables scientists to decode entire DNA strings in minutes and allows them to find new cures and better understand and predict disease patterns. What's more, big data analytics allow researchers to monitor and predict the developments of epidemics and disease outbreaks by integrating data from medical records with social media analytics.
- **Improving sports performance:** Most elite sports have now embraced big data analytics. Video analytics track the performance of every player in a football game, and sensor technology in sports equipment such as golf clubs allows you to get feedback (via your smartphone) on your game and how to improve it.
- **Improving science and research:** Science and research is currently being transformed by big data. Experiments with the Large Hadron Collider, for example, generate huge amounts of data. The CERN (European Organization for Nuclear Research) data centre has 65,000 processors to analyse its 30 petabytes (a petabyte is one quadrillion bytes) of data.

- **Optimising machine and device performance:** Big data analytics help machines and devices become smarter and more autonomous. For example, big data tools are used to operate Google's self-driving car.
- **Improving security and law enforcement:** I'm sure you're aware of the revelations that the National Security Agency (NSA) in the US uses big data analytics to foil terrorist plots (and maybe spy on us!). Others use big data techniques to detect and prevent cyber attacks and police forces use big data tools to catch criminals.
- **Improving and optimising cities and countries:** Big data is used to improve many aspects of where we live. For example, it allows cities to optimise traffic flows based on real-time traffic information as well as social media and weather data.
- **Optimising financial trading:** High-Frequency Trading (HFT) is an area where big data finds a lot of use today by using big data algorithms to make trading decisions.

Using big data in small businesses

Big data might seem like it's something that only big business can make use of. When people first hear that massive volumes of information are being used to fight terrorism, cure cancer or predict the spread of Ebola, it sounds expensive, difficult and time-consuming. But that doesn't have to be the case.

Huge datasets on everything from demographics to weather and consumer spending habits are freely available online for small businesses to use. Plus, the basic tools to make sense of the data are also free and becoming increasingly simple for anyone to use. For example, if you're using Google's AdWords to track what

your customers are searching for online, you're engaging in big data analysis, even if you don't know it.

In many ways, big data is much better suited to small businesses than to big corporations – smaller companies tend to be more agile and able to act on the insights that data provides in a more timely fashion. In the end, even the most impressive data set and the most potent insights are worthless if you don't act on them.

Plenty of small businesses are already using big data to better understand and target customers. Retailers can predict what products will sell, car insurance companies can understand how well their customers actually drive and detect potential fraud and takeaway companies can tailor their services to meet local customer preferences and demand. Social media has become a particularly valuable source of data for understanding customers, trends and markets.

Big data can also help improve business processes. Retailers are able to optimise their stock levels based on what's trending on social media, what people are searching for on the web or even weather forecasts. Supply chains can be optimised so that delivery drivers use less gas and reach customers faster. And you can use data to understand and improve staff engagement or improve your hiring process.

There's more detail on the many big data uses in Chapter 3 – and there are examples dotted throughout the book. Just look out for the Example icon.

Too often I see big data analysis being done in an infrequent, unstructured or ad hoc way (and that's in businesses of all sizes). You really need an underlying strategy in order to get the most out of big data, and there's more on that in Chapter 10. Without an underlying strategy, you may stumble across the odd valuable insight, but with proper planning and preparation, those insights are more frequent and more useful.

Understanding Big Data in More Detail

The first thing to understand is that data in itself isn't a new business phenomenon. Business data is as old as, well, business itself. Just think of sales and financial ledgers or, in more recent history, customer databases. It's specifically *big* data that's the new phenomenon. But, as I mention at the start of the chapter, big data isn't just about how *big* it is. In fact, volume is just one of the key defining factors of big data.

In practice, some of the data you use in your business may not exactly qualify as big data (as defined by the four Vs I explain in the next section), and that's fine. If the best data for you isn't strictly big data, don't lose any sleep over it. So long as you're using data in a strategic way to meet your goals and grow your business, that's all that really matters – not what it's called.

Breaking big data down into four Vs

To understand big data, and what separates it from normal data, you need to understand four main factors, which all handily start with a V. It's these Vs that define what's really special about big data, why it's different to regular data and why it's so transformative for businesses. You can find more information on the Vs in Chapter 2.

The four Vs are:

- **Volume** refers to the vast amounts of data generated every second.
- **Velocity** refers to the speed at which new data is generated and the speed at which data moves around.
- **Variety** refers to the different types of data you can now use.
- **Veracity** refers to the trustworthiness of the data.

I'd include a fifth V that's perhaps more important than all the others: value. It's all well and good collecting vast volumes of data or accessing a wide variety of data, but if you can't turn that data into value (which in the case of business means *growth*) then it's useless.

Why big data is so big right now

I think there are three main reasons why big data is in the news so much these days:

- Big data has incredibly powerful predictive capabilities.
- Big data helps you make much smarter decisions.
- Big data challenges traditional notions of causality.

I look at each of these reasons in Chapter 2.

Another exciting aspect of big data is that it's only going to get bigger and more widely used. As the tools to collect and analyse data become less and less expensive and more and more accessible, we'll develop more and more uses for it – everything from smart yoga mats (no, really) to better healthcare tools and a more effective police force.

Whether you're all for the benefits big data can bring or worried about Big Brother watching everything you do, it's important to be aware of the power of big data. If you live in the modern world, it's not something you can escape. And if you're in business, it's something you should be positively embracing. Having said that, there are some ethical and moral points of debate around big data, such as who owns your data and the need for companies to be more transparent with their customers about what data they collect and why. I explore these issues in Chapter 2.

Turning data into big data

It may seem like big data has exploded onto the business scene out of nowhere. But in fact it's been a more like a gradual evolution: from dusty archive rooms to the microfiche to databases and on to data centres. I think it's part of human nature to want to continually gather information and make sense of what's going on around us – we've just developed sleeker technology for it over the years.

It’s fair to say, however, that the pace of this development has ramped up enormously since the invention of digital storage and the Internet. In particular, three technological advances came together to create the perfect conditions for big data: huge advances in storage capacity, faster networks and more powerful analytic technology.

There’s more detail on the technology changes that underpin big data in Chapter 6. But here’s a short overview of these three advances.

Data storage used to be limited to mainframe computers, hard disks or company servers. Now you have connected computers and servers that give you more storage capacity than ever before. Large amounts of data can be broken up into smaller chunks and stored across a range of machines in different locations. This is called *distributed computing.*

Distributed computing gives you greater storage capacity, but it also allows you to connect data faster than ever before. With data being spread across many different locations, you need to be able to bring that data back together quickly. This is where faster networks come into play.

These massive increases in storage and computing power make number crunching possible on a very large scale. Without faster networks that connect data sets together for analysis, big data just wouldn’t be possible for the average business. Now you can break up the analysis of data into manageable chunks, meaning that

no one machine has to bear the whole load. This makes analysis faster and far more efficient – and cheaper.

For the first time you're able to analyse large, complex and messy data sets (which previously would have been far too big to store). And thanks to new analytic programs, the data can be in just about any form – structured or unstructured, messy or neat, text, audio, video, sensor, images. We now have the ability to extract insights from almost any type of data.

Data, Data Everywhere

It used to be that data would fit neatly into tables, spreadsheets and databases: think of data like sales figures, customer records, wholesale prices and so on. But now you can look at all sorts of data – including emails, Facebook posts, photos, blog comments and voice recordings – and extract meaning.

With big data technology you can now harness different types of data and bring them together with more traditional, structured data. It's this ability to analyse and use a wide variety of data that's especially exciting to me, as it means you can now extract more business-critical insights than ever before.

In this section I give an overview of the different types of data, but you can find more detail (and some great examples) in Chapters 4 and 5.

Discovering structured, unstructured, internal and external data

There are two main types of data: structured and unstructured.

Structured data refers to any data or information located in a fixed field within a defined record or file, usually in a database or spreadsheet. Examples include sales and transaction records, financial data, website hits and customer details.

Structured data has three main things going for it: it's usually cheap to use, it's easy to store, and it's easy to mine for information. But, on the downside, it represents only a small proportion of all the data available these days – as the digital traces you leave behind get bigger and bigger, only a small amount of this data is structured in format.

Another downside is that structured data is simply less rich in insights than unstructured data, meaning it can be more difficult (maybe even impossible) to really understand what's going on if you're using only structured data. For best results, structured data often needs to be paired with other data to get a fuller picture. For example, structured data can tell you that hits on your website increased 20 per cent last month, but you need other forms of data to explore why that happened.

Unstructured data is basically all the data that doesn't fit neatly into traditional formats or databases. Examples includes email conversations,

social media posts, video content, photos and voice recordings. *Semi-structured data* is a cross between unstructured and structured data. It may have some structure that can be used for analysis (perhaps tags such as date or author, for example) but it lacks the strict structure of databases or spreadsheets.

Unstructured and semi-structured data tends to be much more difficult to store (not least because so darn much of it is created every day). Now, thanks to massive increases in storage capacities and the ability to tag and categorise this data, as well as huge leaps in analytical technologies, you can finally make use of this data.

The advantages of unstructured and semi-structured data are that there's absolutely loads of it (it accounts for around 80 per cent of all business-relevant data being generated today), and it providers a richer picture than structured data. However, it's harder to store and more difficult to analyse, which makes it more expensive to work with.

Unstructured/semi-structured data is undoubtedly seen as the sexier, more exciting kind of data – as a result, people often make a beeline for it and neglect poor old structured data. But there's value in both. Both can offer up interesting and useful insights that can help your business. In fact, the real value often comes in combining structured and unstructured data to get a really rich picture of what's going on.

Internal data accounts for everything your business currently has or could access. This may be structured in format (for example, a customer

database or transactional records), or it could be unstructured (conversational data from customer service calls or feedback from employee surveys, for instance). Many people think internal data isn't very exciting, but it can provide a wealth of information.

The beauty of internal data is that it's cheap (or maybe even free) and, as you own the data, there are no access issues to deal with. But, the downsides include having to maintain and secure the data (especially if it includes personal data). You may also find that internal data on its own doesn't provide enough information to meet your strategic goals and you may need to supplement it with external data.

External data is all the information that exists outside your business, whether it's publically available or privately held by a third party. It can also be structured or unstructured in format. Examples include social media data, census data and weather data.

External data is powerful because it gives you access to information that's often more up to date and richer than any information you could gather yourself. And, as it's someone else's data, you have the added bonus of not worrying about the security and data protection issues. But, the obvious downside is that you don't own the data, and you usually have to pay for access (although not always – check out Chapter 15 for some great free data sources).

When it comes down to it, no type of data is really better than any other type. What's best for one business may not be best for yours. The key is to start with a strong data strategy and let that strategy guide you to the best data for you, whether it's structured, unstructured, internal or external or a combination.

Getting acquainted with new types of data

You leave more and more digital traces of your activities than ever before. If you think about what you've done today so far, most of those activities have left some digital trace (data) that can and is being collected and analysed. Some of the data you can now collect is new; some has been around a while but we've only just found ways to really analyse it.

Some of the exciting new types of data include:

- **Activity data:** This is the computer record of human actions or activities that occur online or in the offline physical world.
- **Conversation data:** Conversations are increasingly leaving a digital trail behind – whether it's an SMS message, email, blog or social media post (which are all forms of text data) or as an audio recording of a telephone, teleconference or Skype call.
- **Photo and video image data:** Digital cameras and, more recently, smartphones have resulted in an explosion of this type of data, largely driven by ever-increasing connectivity and a desire to share every aspect of our lives on social media platforms!

- **Sensor data:** Sensors are increasingly being built into products, from phones and cars to golf clubs, all of which results in a vast amount of new data.
- **The Internet of Things (IoT):** Related to sensors and increasing connectivity, the IoT is all about objects being manufactured with embedded sensors and the ability of those objects to communicate with each other.

Making Key Big Data Decisions

There are three keys areas of decision making that relate to big data: one is pulling out insights from the data and using that information to guide your decision making, another is deciding how to build your big data skills and competencies, and the final aspect relates to infrastructure decisions. I look at each in turn in the next sections.

Understanding the value of insights

In today's competitive business world, success often comes down to a company's ability to learn faster than the competition and act on what they learn faster than the competition. The process of turning data into insights and actionable knowledge is the key to that success.

No matter how much data you gather, it's worth very little unless you can turn it into insights and actionable knowledge. Data-based insights help you to make better business decisions. Crucially, you then need to *act* on those decisions. It's this action that creates the fifth V of big data: value.

A key part of this process is making sure the *right information* is delivered to the *right people* at the *right time*. In order to aid decision making and ensure the necessary action is taken, insights need to be presented in a clear, concise and interesting way. People are less likely to act if they have to work hard to understand what the data is telling them.

Data and insights can also feed into the machines in your company, as well as your people. This applies to any machine or technology that's a key part of how the business operates on a day-to-day basis, such as stock control systems or machinery on a production line. Connecting data and machines allows businesses to increase efficiency, improve product quality, cut costs and much more. Processes and systems can also be connected with data, so that you can improve how you do things based on what the data shows.

There's more on focusing on insights and feeding data to your people, machines and systems in Chapter 7.

Building big data skills and competencies

There's currently a skill shortage in big data, meaning there's more demand for big data experts than there are available experts. This can make it hard for smaller businesses to recruit good data staff. But there are alternatives to hiring in-house staff. You can try training up your existing people, working with external data providers (of which there are now many, big and small) and partnering with other organisations, such as universities.

Whether you want to hire new people or boost your existing skills, I think there are six key skills required to successfully use big data in business:

- Analysing data
- Being creative
- Applying mathematics and statistics
- Understanding computer science
- Grasping the business side of things
- Communicating insights

There's more on these skills in Chapter 8.

I think developing your existing staff is a brilliant place to start. So, as a first step, think about whether your existing people have the potential to meet some or all of these needs (with a little extra training in all likelihood). There are some brilliant big data-related courses out there, many of them online and many of them free. I list my favourites in Chapter 8.

If data is going to be a key part of your business, then it's a good idea to consider hiring a data scientist to work in-house. The six skills I list are a good starting point when you're searching for the right person, and I also list some helpful recruitment questions in Chapter 8. If you don't have any tech experience at all then recruiting in the tech field can seem daunting – with these questions and by focusing on the core skills, you'll be able to assess candidates with confidence.

In a competitive recruitment environment where demand outstrips supply, you may need to get a little creative when it comes to tapping into big data skills. It's all about finding creative ways to pull the necessary skills together in whatever way works for you. For example, you may find someone with statistical and analytical skills who falls short on business insights, but your own people could help supplement those skills.

Getting the infrastructure in place

Your ultimate goal is to gather insights which will lead to better decision making and improved business performance. In order to do this, you'll need to invest in some tools or services.

By *infrastructure* I mean the software and hardware that will enable you to turn big data into insights. There are four key elements to a big data infrastructure:

- **Data collection:** If you need to source new data, this may require new infrastructure investments, such as sensors, cameras, or systems to collect text or audio data. For example, if you want to collect machine data from your factory operations or vehicles, you'll need to invest in sensors to collect the data.
- **Data storage:** You need to think about where to store all this data. The main storage options include: data warehouses, data lakes, distributed or cloud-based storage systems, company servers and hard disks.

- **Data analysis/processing:** Now you have your data, the next step is to sort and tidy up the data and analyse it to extract information.
- **Data visualisation/communication:** If the key insights aren't clearly presented, they won't result in action. For most small businesses looking to improve their decision making, simple graphics or visualisation platforms are more than enough to present insights from data.

I explore the main options for each element in Chapter 9, along with some of the most commonly used software packages.

The first step is to assess your existing infrastructure so, for each of these four elements, you need to consider what related technology or resources you already have in-house and how they might need to be improved or supplemented. For example, you may already be collecting useful customer data through your website or customer service department but don't yet have the analytics in place to work with that data.

If you're accessing someone else's data (using Facebook or Twitter, for instance), then the data capture, storage and processing elements may not apply to you – or they may apply to a lesser degree (you may want to partner someone else's data with some of your own internal data). In the last few years many businesses have sprung up offering cloud-based big data services to help other companies and organisations solve their data dilemmas. This type of big data as a service can cut costs because you pay only for the data you use.