

## 1.0. Chapter Overview - Introduction to Big Data

This chapter traces the evolution of Big Data, introduces the value creation processes and examines the concept of Big Data analytics. The chapter also introduces the relevance of privacy and ethics in Big Data context.

1. History and Evolution
2. Big Data in Business and Commerce
3. Privacy and Ethics.

### Quote

“An unsophisticated forecast uses statistics as a drunken man uses lamp-posts -- for support rather than for illumination. ”

Andrew Lang

---

### Chapter Navigation

[Introduction to Big Data](#)

[1.0 Chapter Overview](#)

[1.1 A Historic Look at Big Data](#)

[1.2 Big Data in Business and Commerce](#)

[1.3 An Introduction to Privacy](#)

[1.4 Privacy Policies and Ethics](#)

[1.5 Chapter Summary](#)

## 1.1. A Historic Look at Big Data

### History and Evolution

### Data in Historic Context

### Quote

“A full 90 percent of all the data in the world has been generated over the last two years.”

[SINTEF \(Links to an external site.\)](#), 2013

This quote is striking on two fronts; firstly the sheer volume and secondly the speed at which this trend is increasing due to modern day technologies. Interpretation of data has been around for millennia. Over 7000 years ago, Ancient Mesopotamians tracked basic crops and herd numbers with rudimentary accounting principles. This evolved and throughout time large-scale data sets have been used to record everything from Bubonic plague outbreaks in the 1600s to the first census data collection in the late 1800s.

The Industrial Age was an era of vast societal change, from steam power through to electricity. This led to rapid increases in possible commercial output and, through mechanisation, drastic increases in productivity. The result has been the largest sustained growth of human wealth and capital in the history of humankind. Industrial change demanded an increased labour force, which over time grew more specialised and complex. To coordinate and manage these efforts, business-owners and entrepreneurs developed systems to organise and systematically diffuse responsibility and authority using rules and regulation. This, in effect, was the beginning of the “corporation” as we now know it.

In the 1950s Glickauf and Anderson implemented a payroll system for General Electric (GE) using a digital electronic computer to manage matters of HR. Here, we can see the ushering in of the formative Information Age and its relevance and marriage to the corporation and its role in business (incidentally, the IT consulting industry was also born at this time). The use of technology to manage HR-related information slowly increased during the 1950s to serve many business systems and processes.

**Read: Are you ready for the era of ‘big data’.**

This article from 2011 on Big Data had some interesting predictions. Read this in its historical context and think about what had changed or not gone to plan. Page 6 has a nice graphic on the various industries and their prospective impact.

[Read here. \(Links to an external site.\)](#)

## Evolution of Data

The 1950s also saw the creation of the US National Security Agency (NSA), which specialised in the promotion and furthering of cryptology as a science. Spurred on by the Cold War, the world saw a superabundance of information collection and procurement, rapidly intensifying the importance and prevalence of data and record-keeping. Toward the end of the 1960s, Kleinrock, Roberts, Kahn and Cerf invented a wide-area digital communication network, ARPANET, and this was what we would now consider the beginning of the Network Era and Information Age. This was later popularised by the creation in 1989 of the “world wide web” by Tim Berners-Lee, a British computer scientist.

From the 1990s onwards, data was being created and shared across networks at an amazing rate. Devices became more available and information networking moved from corporate spaces and applications into domestic living rooms and home offices. Personal computing flourished, and in 1995 computing and informatics reached a zenith: the first supercomputer was built. This machine was able to perform as much work in one second as a single calculator being pressed by a lone person for an odd 30,000 years.

## Mini-exercise: Data Timeline

[Data Timeline. \(Links to an external site.\)](#)

Find and locate the following from the link above:

- One key thinker or theorist;
- One key theory/process.

Summarise what their contributions to data science was or has been.

Try to choose something from this timeline that is unfamiliar to you, or a date or concept that piques your interest. Try to consider also the historical significance of this across the many domains here: mathematics, systems, visualisation, computer science.

## The concept of Big Data

### From Data to Big Data

Although the notion of big data was being used in the late 1990s to describe large data sets and volumes, it was only in 2005 that Roger Mougalaas used the term “big data” as we now know it, to discuss notions of impossibility in terms of data management, intelligence and analytical framing. In 2005, Hadoop (more on this later) was also created by Yahoo!, with the intention of mapping and indexing the entire digital web. As these tasks were under way, the rise of social networking sites took flight, and the emergence of Web 2.0 meant that increasing amounts of government, commercial and public data was being created daily.

For example, “Move fast and break things” was Facebook’s original corporate mantra. This was famously changed in 2014 to “Move fast with stable infra”.

## Web 2.0

Refers to a massive shift in Internet applications coinciding with interactivity and social networking platform proliferation: an emphasis on user-generated content, interoperability and usability. Big data refers to the idea that society can do things with a large body of data that weren’t possible when working with smaller amounts. The term was originally applied a decade ago to massive datasets from astrophysics, genomics

and internet search engines, and to machine-learning systems (for voice-recognition and translation, for example) that work well only when given lots of data to chew on. Now it refers to the application of data analysis and statistics in new areas, from retailing to human resources. (The Economist, 2014)

Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse. This definition is internationally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data - i.e, we don't define data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase. Also note that the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple petabytes (thousands of terabytes).

## McKinsey

McKinsey's conceptualisation focuses on the subjectivity of the term, indeed one's Big Data could be another's small computational task. Indeed, this is part of the notion of Big Data: it has become so complex and dynamic in order to be able to process, store, analyse and manage with "traditional" data tools. Indeed, we sometimes cannot even begin to ask the right questions to find the correct answers, as the pace of technological change re-frames this notion of informatics and data-physics constantly. What is important here, for this section, is to acknowledge the history and evolution of data as a quantifiable "thing", albeit also not to ignore the scale and volume as dynamic or shifting as another subject of measurement and focus. Big Data should not be thought of as just a description of volume or storage but more as a way to think about issues of usability.

## Big Data Recorded and Mapped

Another timeline, this one specifically on the origins of the Big Data concept. Check the link below.

[Click here. \(Links to an external site.\)](#)

Note the very interesting framing of the origins of the term around issues of storage, capacity, and some quite fatalistic concerns over speed and computational finity.

## Questions - Reflection

1. When you think of the term "Big Data", what connotations does it conjure for you?
2. Are they positive or negatively framed? Why do you think this is?

[Answer here.](#)

Businesses need to consider the following:

- How do the data they capture and the algorithms used, develop and increase the intelligence of organisations?
- Is it possible to go from return on investment to acquisition of intelligence as a measure of effectiveness and efficiency?
- How can efficiency be improved through understanding – reducing costs and increasing revenues and using data to improve the bottom line and ultimately to create value?
- What processes can we use to turn our information into useful insights, and how do these processes influence decision-making within a business or organisation?

### Watch: Global Leaders Faculty

Mr. Rich Caccappolo, COO of MailOnline, explores the purpose of collecting data.

---

## Chapter Navigation

### Introduction to Big Data

#### [1.0 Chapter Overview](#)

#### [1.1 A Historic Look at Big Data](#)

#### [1.2 Big Data in Business and Commerce](#)

#### [1.3 An Introduction to Privacy](#)

#### [1.4 Privacy Policies and Ethics](#)

#### [1.5 Chapter Summary](#)

## 1.2. Big Data in Business and Commerce

### The Role of Data for Commerce

We are building companies that are smarter and faster-reacting than the humans that run them. Minelli, Chambers and Dhiraj (2012) note the three-pronged perfect storm that allowed Big Data to have its transactional revolution for businesses:

| A computing perfect storm                                              | A data perfect storm                                                           | A convergence perfect storm                    |
|------------------------------------------------------------------------|--------------------------------------------------------------------------------|------------------------------------------------|
| The rise and prevalence of social networks, cloud computing and mobile | The sheer volume, speed and variety of data that is now able to be transferred | The merging of suitcases such as traditional s |

| A computing perfect storm                                                  | A data perfect storm      | A convergence perfect storm                                                                 |
|----------------------------------------------------------------------------|---------------------------|---------------------------------------------------------------------------------------------|
| technology that is cheaper and more accessible than ever for business use. | and shared by businesses. | technologies, to open up data as a commodity and provide new IT alternatives and practices. |

## The Evolution of Business Data Systems

Aside from changes in the physical hardware and software technology, there has been a massive change in the evolution of data systems themselves (or, the processes we implement to solve business issues using data).

Misha notes the following three key stages of data in business:

| Early Days: Dependent                                                                                                                                                  | More Recent: Independent                                                                                                                                 | Big Data Era: Interactive                                                                  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Data systems were relatively new and users didn't know exactly what they wanted. The IT industry simply assumed and adopted a "build it and they shall come" approach. | Users now understood the analytical platform and worked together with IT to define the business needs and approach for deriving insights for their firm. | The interactional stage involves multiple companies, creating collaboration beyond itself. |

## Evolving Concept of Big Data

| Volume                                                                                                                                                                                                                                                                                                                                                                       | Velocity                                                                                                                                                                                                                                                                           | Variety                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| At the rate data is now being created, the amount of data will double every two years. By 2020, the world will generate 50 times that amount of data, according to a 2011 study by IDC. The sheer volume is enormous, and a very large contributor to the ever-expanding digital universe is the "Internet of Things," with all kinds of devices creating data every second. | Velocity is the speed at which data is created, stored, analysed, and visualised. In the Big Data age, data is created in real time or near-real time. With the availability of Internet-connected devices, wireless or wired, machines can pass on data the moment it is created. | In the past, all data was structured that fit neatly into columns and rows, but those days are over. Now, data is unstructured and comes in many different forms: text, structured, semi-structured, unstructured, and even multimedia. |

## Three to Four V's

It is common to see Big Data explained according to three Vs: velocity, variety, and volume. However, Rijmenam's Veracity, Variability, Visualisation, and Value (2014) enables better conceptualisation and understanding of the impact and implications of Big Data on business strategy.

Image: The Four V's of Big Data

IBM data scientists break big data into four dimensions: volume, variety, velocity and veracity.

| Term          | Description                                                                                                                                                                                                                                                                                                                                                   |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Veracity      | The generation and collection of a lot of data at high speed is worthless if the data is incorrect. Incorrect data can cause significant problems for organisations, as well as for the analysis. If you want your organisation to become information-centric, you need to ensure that the data and the analysis are correct.                                 |
| Variability   | Variability is often confused with variety. If a bakery sells 10 different types of bread, that is variety. If, however, the same type of bread tastes and smells different every day, that is variability. Variability is thus very relevant in performing sentiment analysis. Variability is the definition is changing (rapidly).                          |
| Visualisation | This means making a vast amount of data comprehensible in a manner that is easy to understand. Visualisations, of course, are not ordinary graphs or pie charts. They are complex graphs that can include many variables of data while remaining readable and understandable.                                                                                 |
| Value         | (Raw) data in itself is not valuable at all. The value is in the analyses performed on the data. It is how the data is turned into information and then, how this knowledge is put into use. The value is in how organisations use this data to create information-focused commercial value and their decision making on insights derived from data analyses. |

## Big Data For Business

IBMs 9 levers of value creation using Big Data were developed using surveys from across the industry. They acknowledge specifically how organisations derive value through insights embedded in their data collection and management practices. Many of these can be linked nicely to the 7Vs of Big Data, as previously discussed.

### Watch: Levers in Analytics

Please watch this short video and make note of the various “levers” organisations were found to have utilised in regards to their analytics.

### Exercise: Effective Use of Big Data

Select one of the following case studies and diagnose which levers were most employed here in their effective use of Big Data to create organisational and/or commercial value:

- [Disney's guest experiences \(Links to an external site.\)](#)
- [Nestle's supply chain management \(Links to an external site.\)](#)
- [The University of Konstanz \(Links to an external site.\)](#)

[Answer here.](#)

---

## Chapter Navigation

### Introduction to Big Data

#### [1.0 Chapter Overview](#)

#### [1.1 A Historic Look at Big Data](#)

#### [1.2 Big Data in Business and Commerce](#)

#### [1.3 An Introduction to Privacy](#)

#### [1.4 Privacy Policies and Ethics](#)

#### [1.5 Chapter Summary](#)

## 1.3. An Introduction to Privacy

### Privacy and Ethics

#### The Privacy Landscape

How did we get here?

As database marketing gave way to Customer Relationship Management (CRM) and one-to-one marketing, the focus shifted from segmentation to personalising relationships: the more you knew about a consumer, the better you could meet their needs.

There are many reasons why we, as consumers, are perfectly willing to provide personal data (including specifying preferences) in doing business:

- Used responsibly, personal information can enhance our exchanges. Indeed, we are often frustrated by dealing with companies that don't "know" us as individuals, that don't combine our information across products, across organisational boundaries, and across channels, and that don't remember anything about the last interaction we had with them.
- When attempting to resolve issues, we tire quickly of having to repeat ourselves and re-create, yet again, the details of our relationship.

#### The Privacy Landscape

Minelli (2012) has outlined the current data privacy landscape across four key dimensions: businesses, criminals, consumers and litigators.

|             |                                                                                                                                                                                                                                   |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Business    | <ul style="list-style-type: none"> <li>• Increased need to leverage personally identifiable and sensitive information for competitive advantage.</li> <li>• Significant investment in data sources and data analytics.</li> </ul> |
| Criminals   | <ul style="list-style-type: none"> <li>• The dramatic surge in identity theft.</li> <li>• Sophisticated technology to exploit data-security vulnerabilities.</li> </ul>                                                           |
| Consumers   | Increased awareness and concern about collection, use and disclosure of their personal information.                                                                                                                               |
| Legislators | <ul style="list-style-type: none"> <li>• Responding to consumer concern by restricting access to and use of personal information.</li> <li>• Significant impact and restrictions for business.</li> </ul>                         |

## The Effects of Big Data on Privacy

Big Data is considered to be a threat to privacy for several reasons:

- Big Data is permanently available.
- Big Data involves large volumes of data.
- Big Data on customers is collected invisibly.
- There is no good assessment of the privacy sensitivity of data.
- Due to the large volume of data, it is no longer accessible and understandable by customers.
- Data from multiple sources are being fused.
- Customers perceive a lack of control on Big Data collection.

## The Security Database

The promise of Big Data is accompanied by perils—in terms of control, privacy, security, reputation, and social and economic disruption. How will we manage these trade-offs individually and in business, government, and society?

Debates have been going on for years and will continue as we work out the differences between “rights”: the individual’s rights to determine what personal information they are willing to give in exchange for services, versus the service provider’s rights to determine how to continue to provide these services. When discussing services, we can think about the many different applications, software, online platforms, hardware and technologies used where a service is being provided by an organisation.

We can view this holistically as a mutual agreement of exchange of value. In the case of online digital product and service agreements, the individual has only one immediate binary choice: accept the terms as stated or do not use the service. There is very rarely negotiation of these terms of setting of conditional exchanges. The fact is that most individuals do not comprehend the computational power and possibilities that Big Data

brings. Some think it would be impossible for someone to glean deep insight into their private lives.

However, the pervasiveness of Big Data makes it harder to keep this information private.

## What Information is Collected?

When we say “pervasive” we mean quite wide-ranging. The following table outlines the full scope of data and information that could be (we can debate the should be later!) collected and stored from technology and digital interactions.

| Personally Identifiable Information (PII)                                                                                                                                                                                                                                               | Sensitive Information                                                                                                                                                                                                                                                                   | Other Information                                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Any information that directly or indirectly identifies a person.                                                                                                                                                                                                                        | Any information that, if disclosed without authorisation, could be embarrassing or detrimental to the individual.                                                                                                                                                                       | Any other non-identifiable information about an individual PII.                                               |
| <ul style="list-style-type: none"><li>• Name</li><li>• Postal address</li><li>• Email address</li><li>• Telephone number</li><li>• Medicare number</li><li>• Driver’s license number</li><li>• Bank/Financial number</li><li>• Credit or debit card number</li><li>• Postcode</li></ul> | <ul style="list-style-type: none"><li>• Race/ethnicity</li><li>• Political opinions</li><li>• Religious beliefs</li><li>• Trade union memberships</li><li>• Health/medical info</li><li>• Marital status/sexual life</li><li>• Age</li><li>• Gender</li><li>• Criminal record</li></ul> | <ul style="list-style-type: none"><li>• Preferences</li><li>• Cookie ID</li><li>• Static IP address</li></ul> |

While data volumes are clearly greater today, the great data grab has been going on for decades. One primary means of past data exchange was via the warranty card that came with products. Under the guise of registering a product to ensure warranty rights, companies gathered personal information about individuals. These cards were not managed by the companies themselves but were outsourced to processing centres, who then had access to data across products. They gladly bundled and sold this data to other companies.

## Metadata

What exactly is Metadata and how does it relate to Big Data?

The simplest definition of metadata is “structured data about data”. Metadata is information about specific pieces of data, which is used to improve and enhance the use of that piece of data. More simply, metadata provides the context for a given data set.

Metadata is descriptive information about an object or resource, whether it be physical or electronic.

For example, metadata has a long history of use in library catalogue systems. In a library, each piece of data (the individual cards within a library card catalogue) features metadata about where to find the book within the library, who wrote the book, what the book is about, and more. This metadata gives context to the card as a means to identify which books are in a particular library. Today, metadata is commonly used in information technology systems to identify the programs used to compile data, who compiled the data, how to access and alter the data, and much more. Metadata can be generated either “by hand” or derived automatically by using software.

### **Watch: How Google collects and stores metadata**

This video, while critical, takes a nice look at how Google collects and stores a wide variety of metadata. See if you can spot the many instances of automated collection, and the various many industries.

### **Read: The Parable of Google Flu: Traps in Big Data Analysis**

An excellent article looking at the role businesses (here, Google) can play in solving social ills, through technology. This is a well-cited article and nicely explores the role of hubris, transparency, as well as expectations in regards to technological determinism.

[Read here.](#) (Links to an external site.)

### **Image: What Do Companies Do With Data?**

A graph describing the extent of regulation and enforcement of personal data collection. In Canada and the UK regulation and enforcement is considered heavy, whilst China and Brazil experience limited control.

Minelli, M., Chambers, M., & Dhiraj, A. (2012)

So, what can companies do with all the personal data they collect? How do we know that they are doing what they say they will? When exactly are our rights violated? Why should it matter to us? Most companies develop their own privacy policies as a matter of establishing a modicum of “trust” with consumers.

Sensitive data is collected with personal information. In certain countries and markets, additional restrictions apply to the collection, use, or disclosure of sensitive data. Note that many companies have sensitive data policies however these policies may be focused on data that is sensitive to the company itself, and it may, or may not, include reference to personally sensitive data. From the perspective of data management

policies, differentiating and managing such data appropriately is a matter of data classification.

---

## Chapter Navigation

### Introduction to Big Data

#### [1.0 Chapter Overview](#)

#### [1.1 A Historic Look at Big Data](#)

#### [1.2 Big Data in Business and Commerce](#)

#### [1.3 An Introduction to Privacy](#)

#### [1.4 Privacy Policies and Ethics](#)

#### [1.5 Chapter Summary](#)

## 1.4. Privacy Policies and Ethics

### Seven Global Privacy Principles

Beyond data legislation, companies should be aware that the resulting policies from their data usage are also affected by laws. For example, if firms target personal characteristics, such as religion, gender and race, this could result in discrimination. Although the firm could act in accordance with specific privacy legislation, its resulting policies could be in conflict with regulations on discrimination. Hence firms should look beyond the privacy legislation, and should also consider regulation related to specific marketing actions. The next section outlines the seven EU-US Safe Harbour Principles which most companies have integrated into their Big Data policies.

| Policy                   | Description                                                                                                                       |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| 1. Notice (Transparency) | Inform individuals about the purposes for which information is collected.                                                         |
| 2. Choice                | Offer individuals the opportunity to choose (or opt out) whether and how personal information they provide is used or disclosed.  |
| 3. Consent               | Only disclose personal data information to third parties consistent with the principle of choice.                                 |
| 4. Security              | Take responsible measures to protect personal information from loss, misuse, or unauthorized access, disclosure, or modification. |

| Policy            | Description                                                                                                                               |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
|                   | access, disclosure, alteration, and destruction.                                                                                          |
| 5. Data Integrity | Assure the reliability of personal information for its intended use and reasonably ensure information is accurate, complete, and current. |
| 6. Access         | Provide individuals with access to personal information data about them.                                                                  |
| 7. Accountability | A firm must be accountable for following the principles and must include mechanisms for assuring compliance.                              |

## Ethics

At a higher level, one could debate whether firms should mainly focus on legislation and take that as the rule on how far they will go with data collection or whether they should take a broader perspective. The latter leads to a discussion on how organisations make “moral” decisions. We could advocate an approach in which firms are allowed to collect data as long as the law allows them to do so. However, we could also adopt a more ethical perspective that goes beyond laws, where firms consider that they have an ethical responsibility.

Ethical decision making applied to privacy involves four stages:

1. Firms have to recognise that privacy decisions have a moral dimension.
2. Firms then have to form an ethical judgment concerning what ought to be done with regard to privacy.
3. Firms have to establish the moral intention to act in conformity with what they have judged to be the right type of behaviour with regard to privacy issues.
4. Firms have to engage in that behaviour.

## Image: Privacy for Value Creation

An axis representing Thomson Reuters value creation possibilities.

Organisations should strongly consider the ethical and reputational consequences of their Big Data and privacy policies. Specifically, they should adopt stakeholder management and consider reactions from customers, the government (including politicians), and the media. Thomson Reuters outlines the value creation possibilities that can stem from an appropriate and well-managed data program with respect to privacy. By leveraging reputation and legitimacy, firms can deliver behavioural data on their products and services in a manner that meets business and consumer needs for well-defined data types, location, data ownership and lifecycle.

The reputational consideration is highly relevant, especially to the emergent Millennials consumer segment. The 2015 Ranking Digital Rights Corporate Accountability Index evaluates 16 of the world’s most powerful Internet and telecommunications companies

on their public commitments and disclosed policies affecting users' freedom of expression and privacy. Very interesting to see who makes the list!

[View the list here.](#) (Links to an external site.)

## Watch: Big Data is better data

Similar to the above, but far more optimistic! Looks at machine learning and the links to big data – how design can be aided in the future, as well as some exciting “what’s next” futurism!

---

## Chapter Navigation

### [Introduction to Big Data](#)

#### [1.0 Chapter Overview](#)

#### [1.1 A Historic Look at Big Data](#)

#### [1.2 Big Data in Business and Commerce](#)

#### [1.3 An Introduction to Privacy](#)

#### [1.4 Privacy Policies and Ethics](#)

#### [1.5 Chapter Summary](#)

## 1.5. Chapter Summary

In this chapter we considered:

### A history of data

An overview of the nature of data and how we came to the end of 2017 with data so prevalent. We also unpacked the concept of Big Data and its historical and semantic history. Big Data should not be just a description of volume or storage, but framed around issues of usability.

### Data for business

After discussing the notion of Big Data, we looked more closely at the role of data to the business and commercial landscape. The value is in how organisations use this data to create information-focused companies that base their decision making on insights derived from data analysis.

## Analytics and the 7Vs

Rijmenam's Veracity, Variability, Visualisation and Value were added to better conceptualise and understand the impact and implications of Big Data on business strategy. This helps to frame our thinking in terms of strategic implications.

### Exercise

Try briefly answering these review questions:

1. What key historical (perhaps global) events led to the development of data for a monitoring and processing perspective? Why?
2. Why is there so much conjecture around the definition of Big Data?
3. Explain how you see the relationship between ROI and analytics in modern day business? Why is this relationship so important?
4. Why do you think the extra four Vs are a good addition to the earlier three? What do they illuminate and provide in terms of a deeper relevance to business analytics in particular?

[Answer here.](#)

### Extra Resources

| Resource                                                                                   | Description                                                                                                                                                                                                                                                                                                                               |
|--------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                            | Light, easy and fun                                                                                                                                                                                                                                                                                                                       |
| The New York Time's "Eight, No Nine, Problems with Big Data", Marcus and Davis, April 2014 | A nice critical overview of the Big Data Z realism!<br><a href="#">Read here. (Links to an external site.)</a>                                                                                                                                                                                                                            |
| "Big Data is better data", Kenneth Cukier, TedX                                            | Similar to the above, but far more optimistic about machine learning and its links to big data. It can be aided in the future, as well as some futurism!<br><a href="#">Watch here. (Links to an external site.)</a>                                                                                                                      |
| Privacy by Design (2016)                                                                   | Some interesting publications are released on Privacy by Design (PbD). Their approach to protect data is to embed it into the design specifications of business practices, and physical infrastructure. This approach to privacy policies and information management in firms.<br><a href="#">Read here. (Links to an external site.)</a> |
| The Economist (2015), "We'll see you, Anon"                                                | <a href="#">Read here. (Links to an external site.)</a>                                                                                                                                                                                                                                                                                   |

| Resource                                                                                                                                            | Description                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| The Stanford Law Review                                                                                                                             | Looking at the standards around data collection, this is very easy to read, and touches on some of the automation/metadata debates that have been in the media and wider society concerns. Another article by The Stanford Law Review is focused on the policy and nitty gritty)<br><a href="#">Read here. (Links to an external site.)</a> |
| For the academically curious                                                                                                                        |                                                                                                                                                                                                                                                                                                                                             |
| Jacobs, A. (2009). The pathologies of big data. Communications of the ACM, 52(8), 36-44.                                                            | A lively reading on the pathologies of data collection, computational, but rather analytical. Jacobs create a meta-definition of the concept, the shifting nature of historical data analysis, but stick with it!<br><a href="#">Read here. (Links to an external site.)</a>                                                                |
| Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, 36(4), 1165-1188. | An excellent publication-based framework for business intelligence and analytics literature, looking at some of the issues around classification, as well as the evolution and identification of the Big Data corpus.<br><a href="#">Download here. (Links to an external site.)</a>                                                        |

## References

BB (2016) "Safe Harbor Primary Principles", accessed online at: <https://www.bbb.org/council/eusafeharbor/bbb-eu-safe-harbor-dispute-resolution-program/safe-harbor-privacy-principles/> [Feb, 2016]

DublinCore (2016) "DCMI FAQs", accessed online at: <http://dublincore.org/resources/faq/> [Feb, 2016]

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute. Available online at: [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Ohlhorst, F. J. (2012). Big data analytics: turning big data into big money. John Wiley & Sons.

The Economist (2014), "Big Data Explained", accessed online at:  
<http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-10>

Van Rijmenam, M. (2014). Think Bigger: Developing a Successful Big Data Strategy for Your Business. AMACOM Div American Mgmt Assn.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge.

University of Warwick. Centre for Corporate Strategy and Change. (1988). The management of strategic change. A. M. Pettigrew (Ed.). B. Blackwell.

---

## Chapter Navigation

### Introduction to Big Data

#### 1.0 Chapter Overview

#### 1.1 A Historic Look at Big Data

#### 1.2 Big Data in Business and Commerce

#### 1.3 An Introduction to Privacy

#### 1.4 Privacy Policies and Ethics

#### 1.5 Chapter Summary

---

## Chapter 2

### 2.0. Chapter Overview - Big Data in Business

This chapter explores some of the business environments where Big Data is applied, in real time, through examples and the evolution into business. It also looks at current and emerging trends going into 2018.

1. Business and Data.
  2. Applications.
  3. Analytics.
  4. Business Trends.
-

## Chapter Navigation

### Big Data in Business

#### [2.0 Chapter Overview](#)

#### [2.1 Business and Data](#)

#### [2.2 Data Applications](#)

#### [2.3 Analytics and Big Data](#)

#### [2.4 Big Data Evolution and Business Trends](#)

#### [2.5 Chapter Summary](#)

## 2.1. Business and Data

### Origin of Business Data

The origin of data management coincides with the wide adoption of computers. The discipline of statistics existed well before computer science, but computers empowered statisticians to solve a wide variety of practical problems with real life implications, since heavy number crunching and massive storage of data became feasible due to the emergence of modern computing technologies. The invention of database management systems in the 1960s, and relational database management systems in the 1970s, accelerated the pace of this marriage between statistics and computer science.

- In the late 1980s, terms such as “knowledge discovery” and “data mining” emerged.
- In the early 1990s, database industry practitioners started noticing the explosion of business data in the form of Big Data. The official use of the phrase Big Data can be traced back to an article published in the ACM Digital Library in 1997.
- In the late 1990s, the phrase “data science” first appeared to inspire researchers and professionals.

### The Role of Data for Commerce

In the previous chapter, we addressed Minelli, Chambers and Dhiraj’s (2012) three-pronged perfect storm that allowed Big Data to enact its seismic revolution upon businesses:

| A computing perfect storm                                                                                                                         | A data perfect storm                                                                                     | A convergence perfect storm                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| The rise and prevalence of social networks, cloud computing and mobile technology that is cheaper and more accessible than ever for business use. | The sheer volume, speed and variety of data that is now able to be transferred and shared by businesses. | The merging of suitable technologies, to open up new commodity and per |

| A computing perfect storm | A data perfect storm | A convergence pe                  |
|---------------------------|----------------------|-----------------------------------|
|                           |                      | provide new IT alte<br>practices. |

This is especially relevant in our framing of the Big Data industry, as it nicely conceptualises the various key influences on its rise in both relevance and function to businesses. Think closely here about how IT and software, as well as technological and societal acceptance, has allowed for the rise and ubiquity of information and data in the marketplace more widely.

## Big Data Still Unappreciated?

We also explored McKinsey's key definitions and conceptualisation of Big Data, which focused on the volume, rate and pace of data management and collection. This watershed report written in 2011 by the firm also hailed Big Data as the next "frontier" for innovation, competition and productivity. However, managers and business leaders are still in some ways frustrated because their organisations do not give data and information the attention they deserve; most data sits unused, with opportunities to capitalise on this data (think R&D potential, customer insights and transactional or market information) sometimes completely ignored.

### Read: Eight, No Nine, Problems with Big Data

A nice critical overview of the Big Data Zeitgeist and a call to realism!

[Read here. \(Links to an external site.\)](#)

## Not Just Business Assets

What is most illuminating, and perhaps striking, is that bad data management and interpretation has been at the heart of more recent mismanagement and improper business functions, and has even had global ramifications: think about the subprime mortgage meltdown and following market volatility. There are also the more socially relevant effects: hospital and medical diagnosis errors, identity theft and fraud, as well as the effect of contested elections in democracies.

## Out of The Office and Into The Quantified Self

The use of personal data to improve one's health and well-being. The concept of Quantified Self entails tracking exercise, sleep patterns, diet and other factors and making lifestyle changes based on the results. The term "Quantified Self" also refers to an informal group of manufacturers and users of self-tracking instruments that connect through live conferences and online forums. When organisations fail to recognise the importance of data and information, chances are there is a trickle-down effect on consumer awareness (or lack of it!). What impacts the market and economy is also

likely to impact society, and the benefits of this relationship have led to suitable and relevant products, improved services, lower prices for necessities, and accountability and transparency practices.

We must also make note here of the consumer's own willingness to harness and create data. Previously, we looked at the impact of Web 2.0 technologies for networking and sharing of data. The "quantified self" revolution presents itself as a pertinent example of the information age and societal adoption. This relationship between society driving technological change, and the market and organisational impacts, is best explained by the Technology-Impact cycle.

Investopedia (2015).

### Image: The Technology Impact Cycle

This is considered an ongoing cycle which takes many years for a full transference. If we apply this to our previous definitions of Big Data and see that it was first explored as a concept in 1997, and then again in 2005, we can see that organisations and the economy are only now addressing these in business and commercial terms.

Think about the following societal and technological revolutions.

- Have they followed the Technology Impact (TI) cycle here?
- Have they progressed into other sectors? Why or why not?
- Where do you think they originated?

| Societal                                                                                                                                                      | Technological                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"><li>1. A shift to handmade, artisan, crafted products and bespoke services</li><li>2. Recycling and eco-awareness</li></ol> | <ol style="list-style-type: none"><li>1. Wi-Fi technology</li><li>2. Fingerprint and biometric scanning</li></ol> |

---

## Chapter Navigation

[Big Data in Business](#)

[2.0 Chapter Overview](#)

[2.1 Business and Data](#)

[2.2 Data Applications](#)

## 2.2. Data Applications

### Image: Data Applications

#### A Short History

The 1980s was all about ERP and applications designed to make an organisation's processes more effective and efficient. By better managing all of the key functions of a company – accounting, warehouse, inventory and supply chain – ERP allowed the company to focus on what really matters: the information and raw data.

ERP: Enterprise Resource Planning

CRM: Customer Relationship Management

Minelli et. al (2012).

Customer Relationship Management (CRM) systems focus on exactly that: customers. They aim to manage all customer information and capture relevant details for the company which can ultimately build their customer relationships in order to increase customer loyalty and therefore revenue.

During the CRM era of the 1990s, many companies invested heavily in customer-facing technologies which were not overly effective. Management did not fully comprehend that big projects require a coordinated and business-wide transformation of people, process, and technology. All three must be in harmony or the desired transformation is negligible. The arrival of eCommerce and analytics in the 2000s saw this convergence of several trends as discussed earlier — more data and less-expensive, faster hardware. Currently, in this age of Big Data business analytics, we have faster processing at an affordable price that is within reach and available to most businesses.

Reflect and think about how insurance companies, in particular, have evolved with these data system changes. Think about what they are now capable of in terms of not just managing their customers, but also projection and risk assessment, all the way to complex prediction and forecasting of multiple variables.

#### Data in Many Guises

The variety of data sources continues to increase as applications evolve. Internally focused operational systems, such as ERP and CRM applications, were once the major sources of data used in analytic processing. However, in order to increase knowledge and awareness, the complexity of data sources that feed into analytic processes is rapidly growing to include a wider variety of data sources such as:

| Data Source                        | Example                                                                 |
|------------------------------------|-------------------------------------------------------------------------|
| Online and web data                | Clickstream analysis- this is coming up!, social media sites            |
| Location-based and geospatial data | From mobile devices                                                     |
| Data from devices                  | RFID sensors, LIMs and medical devices, telemetry, NFC                  |
| Imaging data                       | Online and mobile video, satellite images, surveillance and observation |
| Supply chain information           | Ordering, vendor catalogs and pricing information                       |
| Primary research methods           | Surveys and experiments                                                 |
| Secondary research                 | Industry and marketplace data                                           |

## Data and Healthcare

The changing health-care landscape is an excellent example of where Big Data advances and analytics can have an immediate beneficial impact. Revolution of the health-care system will come through Big Data-driven decisions and improved insights from a variety of information sources. Important advancements from the management of chronic disease to the delivery of personalised medicine are now possible due to the insights able to be extracted from large and complex data sets being collected in this industry. In addition to saving and improving lives, Big Data has the potential to transform the entire health-care system with objective, data-driven insights. Data here can range from:

- Biological data, including gene expression and large-scale epigenetics
- Special Needs Plans (SNPs) for those with chronic disease or disability and impairment
- Proteomics and the study of proteins and genomes
- Metabolomics and small-molecule profiling
- Gene-sequencing data

This exponential growth in data is further fuelled by the digitisation of patient data. This is now more commonly stored in Electronic Health Records (EHRs) and Health Information Exchanges (HIEs) and supplemented with data from imaging and test results, medical and prescription information, and, increasingly, personal health devices and wearable technology.

An interesting infographic on the sheer numbers of this industry data revolution

[Click here.](#) (Links to an external site.)

And also Google's pioneering and now finalised efforts on flu-mapping

[Click here \(Links to an external site.\)](#)

**Read: Big data: the next frontier for innovation in therapeutics and healthcare.**

A top-line and straightforward overview of the impact of Big Data and informatics on pharmacology. An excellent, if short, overview of the enabling potential of data in this industry.

[Download here. \(Links to an external site.\)](#)

## Medicine

Concurrently, medicine is undergoing a seismic shift with the increasing adoption and application of data and analytics into both laboratory and practicing fields of industry. Now, datasets are available that contain thousands of genes. Typically, medical datasets now contain somewhere between 100,000 to 500,000 genetic variations. Algorithms are developed and uncovered that identify the interactions (say, between environmental factors and diseases) which are assisted by rapid built-in search techniques. Statistical analysis is also available on these datasets: permutation analysis which may require incremental change is now assisted using large-scale computing mechanisms. This mean that it is not impossible to achieve 500,000 genetic variations in a single patient sample for diagnosis or investigation.

Pharmacology is also undergoing a transformation through similar Big Data applications:

Innovation, improving research efficiency, increasing time to clinical trial, while building new tools for physicians, consumers, insurers, and regulators to assist in more specialised drugs and medicine.

One of your readings for this chapter is the McKinsey article on this sector: not to be missed!

## Business Operation Applications

We should also not forget the many various applications within businesses and the function of data and information in regards to internal operations. Such as:

| Operation           | Description                                                                                           |
|---------------------|-------------------------------------------------------------------------------------------------------|
| Procurement         | identifying which suppliers are most cost-effective in delivering products on-time.                   |
| Product development | uncovering product usage insights to speed product development processes and in launch effectiveness. |
| Manufacturing       | highlighting machinery and process variances that might be indicators of quality pro                  |

| Operation                    | Description                                                                                                                                                 |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Distribution                 | quantifying optimal inventory levels and optimising supply-chain activities based on such as weather, holidays, and economic conditions.                    |
| Marketing                    | identifying which promotions and campaigns are most effective in driving customer and sales.                                                                |
| Pricing and yield management | optimising prices for “perishable” goods such as groceries, dynamic pricing such as concert tickets and fashion merchandise.                                |
| Sales                        | optimising sales resource assignments, product mix, commissions modeling, and a                                                                             |
| Store operations             | optimising stock levels, given buying patterns, coupled with local and real-time demand and events data.                                                    |
| Human resources              | identifying the characteristics and behaviours of your most successful and effective how to manage this resource efficiently with variable buying patterns. |

In 2013, pharmaceutical giants, Merck, used analysis to dramatically cut the amount of waste caused by variance in manufacturing environment conditions. It took three months and involved 15 billion calculations on individual production data from 5.5 million vaccine batches. This allowed them to discover the optimum conditions during the fermentation process and should greatly increase their yield, once the FDA has approved the proposed changes to the manufacturing process.

## The Advertising Sector

Traditionally, the advertising industry had quite powerful and useful channels and tools to reach their audiences: media such as billboards, newspapers, radio stations and television. Clients to this industry were overly focused on dominating these channels: trying to reach as many consumers as possible and aiming for volume over targeted selection or distribution of message. As the industry matured and as advertising space became cluttered and at a premium, advertisers were forced to learn more about their audiences. Media research firms stepped into the void to statistically measure channels and record and survey who watched or read broadcasts and at which times. In this way research firms developed segments and profiles which in essence could be targeted. This enables advertisers to best place their ads and target and position their content to the most relevant audience, making “media spend” more data-informed and efficient.

Analytics in this sector also helped advertisers answer the following questions:

- How much do I need to spend?
- How do I allocate that spend across all the marketing communication touch points?
- How do I optimise my advertising effectiveness against my brand equity and ROI in real-time?

Today, agencies are devoted to metrics and real-time analytics to track audiences and convey this information back to clients for more precise indicators of audience engagement.

## Watch: Global Leaders Faculty

Ms Sarah Watson, Chief Strategy Officer of BBH in New York, discusses how to approach advertising with diverse mindsets.

## Read: Big Data's Big Impact on the Future of Advertising

Barns, who is the CEO of Nielsen (the large market survey and insights firm), gives a succinct defence of why it is not about big data but the; right data

[Download here. \(Links to an external site.\)](#)

## Towards Market Automation

Database marketing is concerned with building databases containing information about individuals, using that information to better understand those individuals, and communicating effectively with some of those individuals to drive business value. Historically, database marketing was mostly direct mail. As companies grew and systems proliferated, we ended up with a landscape where you had one system for one product, another system for another product, and then potentially another system for another product, leading to a very silo-ed and disconnected approach to CRM.

In this age of analytics, we now have the ability and systems available to use database marketing for customer acquisition. Companies began to develop technologies to manage and duplicate data from multiple source systems. Primarily, companies started collecting information from customer care centers. The data became richer, creating a better picture of the individual and the mix of products and services purchased. We had much better sets of data around who's contacting whom, and could proactively seek and target these groups.

Now, proactive communication is commonplace, and the arrival of newer and more effective marketing automation solutions accelerates the trend. Marketing automation refers to software platforms designed to automate repetitive tasks in critical areas such as campaign management and allows agencies and in-house marketing teams to reduce lead times, and maximise the efficiency of a targeted campaign.

## Watch: Global Leaders Faculty

Mr John Buchanan, Founder of Buchanan Success Consulting, discusses the use of data in performance management.

---

## Chapter Navigation

### Big Data in Business

#### [2.0 Chapter Overview](#)

#### [2.1 Business and Data](#)

#### [2.2 Data Applications](#)

#### [2.3 Analytics and Big Data](#)

#### [2.4 Big Data Evolution and Business Trends](#)

#### [2.5 Chapter Summary](#)

## 2.3. Analytics and Big Data

### Analytics

#### Social Media Analytics

More and more people are using social media. This generates an enormous amount of data. Data scientists are naturally attracted to these emerging types of data sets. Social media refers to websites where users can post their own content and share it with friends and others. Depending on their focus, social media sites promote different types of interests.

For example, Facebook offers a forum for building informal and personal relationships, compared with a professional networking tool like LinkedIn. In addition to its size qualifying as big data, another unique value of social media data lies in the data about data, or metadata, it carries. For example, a post on Facebook can accompany location information as well as timestamps. With these kinds of unstructured but very rich data sets, a lot of useful insights can be derived about a person who is posting and consuming information. Another example is that IBM has a product called Personality Insights which offers a profiling service for companies.

#### Watch: Global Leaders Faculty

Mr Dominic Barton, Managing Director of McKinsey & Co covers analytics and performance.

#### Disease Control

One area of social media analytics applications is disease control. University of Pennsylvania researchers conducted a study on a predictive relationship between Twitter post content and heart disease. Emotional factors are linked to heart disease.

The study identified indicators of emotional distress expressed in words and correlated them to the occurrences of heart disease. Their study used linguistic analysis techniques as well as various Big Data analytics techniques to reveal key words of emotion such as hate to be strongly correlated to the incidence of heart disease. Conversely, positive words such as wonderful showed the opposite correlation. The Twitter data they collected consisted of tweets posted by 88 percent of the people from countries in 2009 and 2010. This ample data set provided much stronger evidence of correlation than what they could provide through conventional surveys of subjects.

[Read more here. \(Links to an external site.\)](#)

## SAP Hana

Please read the following extract from an interview with Steve Lucas, SAP Hana Executive and VP on Big Data Technologies.

“You don’t have to guess. You can look at actual data, from real customers. You can extract and analyse every policy they’ve ever held. The answers to your questions are buried in this kind of massive mound of data — potentially petabytes worth of data if you consider all of your insurance customers across the lifespan of their policies. It’s unbelievable how much information exists. But now you’ve got to go from the level of petabytes and terabytes down to the level of a byte. That’s a very complex process. But today you can do it — you can compare one individual to all the other people in an age bracket and perform an analysis, in real time. That’s pretty powerful stuff. Imagine if a customer service rep had access to that kind of information in real time. Think of all the opportunities and advantages there would be, for the company and for the customer.”

Here’s another example: You go into a store to buy a pair of pants. You take the pants up to the cash register and the clerk asks you if you would like to save 10 percent off your purchase by signing up for the store’s credit card. 99.9 percent of the time, you’re going to say “no”. But now let’s imagine if the store could automatically look at all of my past purchases and see what other items I bought when I came in to buy a pair of pants — and then offer me 50 percent off a similar purchase? Now that would be relevant to me. The store isn’t offering me another lame credit card — it’s offering me something that I probably want, at an attractive price.”

- Explain and dot point the benefits to both business and consumer for this transaction.
- What is the value of real-time analytics to this industry in particular?
- What benefit is there to customers? Why would they choose to shop at this store?

## The Emergent Big Data Universe

One of the most significant benefits of the emergence of Big Data analytics for businesses is that they are not constrained by predefined sets of questions or queries. With traditional analytics, the scope of questions they could ask the database is

(relatively) extremely small. With Big Data analytics, this universe is vast. Businesses and organisations can define new variables in the moment, dynamically, or “on the fly”. This is a very different scenario from the previous approach, in which our ability to ask questions was limited and constrained by programming and initial planning.

**Why is the ability to define new variables so critically important?**

The answer is straightforward. In the marketplace, you don’t always know what you are looking for. You can’t possibly know in advance which questions you’ll need to ask to find a solution. The following covers some emergent applications of Big Data that are opening up due to this very premise. The questions have not been asked in advance, yet we are seeing results and uses of analytics that are causing disruption and upheaval in certain industries.

## WEB 3.0 and Beyond

Web 2.0, as we know it, is on its way out the door. For many, Web 2.0 is characterised mainly by the ability of users to share information quickly with others, which developed into the phenomenon that we call “social media”. Web 3.0 is being referred to by experts as the semantic web; semantic meaning data driven. The data will come from the user and the web will essentially adjust to meet the needs of the user.

For example, if you do a lot of searching for “design blogs”, you’ll receive more advertisements related to design. Also, when you search for other things, for example, “computers”, the web will keep in mind that you often search for design and may pull up search queries that combine “design” and “computers”. As a result of the growing popularity and use of social media around the world and across nearly every demographic, the amount of user-generated content — or Big Data — created is immense, and continues growing exponentially. Millions of status updates, blog posts, photographs, and videos are shared every second. Successful organisations will not only need to identify the information relevant to their company and products but also be able to dissect it, make sense of it, and respond to it, in real time and on a continuous basis, drawing business intelligence, or insights, that help predict likely future customer behaviour.

## Digital Marketing

Big Data on the web will completely transform a company’s ability to understand the effectiveness of its marketing and hold its people accountable for the millions of dollars that they spend. It will also transform a company’s ability to understand how its competitors are behaving. Data-driven innovations have led to the emergence, and importance, of:

|                      |                                                                                                                                                              |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Clickstream Analysis | Collection and analysis of the “online fingerprint”: sophisticated tracking on w in close to real-time, and embedded in many “pushed” marketing content item |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                            |                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Affiliate marketing                        | Word-of-mouth marketing has been the most powerful form, long before the Internet. The complex digital architecture is what can link consumers via referrals to likes and shares.                                                                                                                                                                                                                                        |
| Multi-platform management                  | Coordination of campaigns and content across multiple-platforms, with monitoring able to be strategically aligned.                                                                                                                                                                                                                                                                                                       |
| Social media and influencer identification | Smart software is emerging that crunches large social datasets to define things like the value of a post and likelihood of response to a piece of online content. To some brands, influence is measured purely by reach, and to others, true influence is more of a function of the thoughtfulness of posts showing a real understanding of a given topic and yielding influence via social engagement or conversations. |
| Cross-channel lifecycle marketing          | This is the increasingly prevalent capturing of customer permissions, contact information and preferences across multiple channels. It also requires marketers to have the technology and marketing and customer information systems for recording and automation for use throughout the customer lifecycle.                                                                                                             |

**Watch: The curly fry conundrum: why social media “likes” say more than you might think.**

Watch the TED talk and reflect on the following questions:

- Did anything shock you about the pervasiveness of the data collection tools and methods?
- Can you personally remember a time on social media where you felt as if your information had been adopted for commercial reasons?
- Do you think there are any risks with these increasingly commonplace approaches?
- Why would online businesses (think e-tailers) be highly interested in advanced digital marketing practices such as this?
- How do you think a start-up company or enterprise might find these tactics appealing? What information would they be most interested in?

[Answer here.](#)

## Mobile

There is no need to explain the prevalence and ubiquity of mobile devices but this ties in nicely to our discussions around Web 3.0 and the semantic web. The number of mobile users is expected to increase in the coming years, with a greater dependency on mobile in general. As of 2016, 61% of web traffic came from wireless devices as opposed to desktops, contributing heavily to the growth of Big Data. Companies are attempting to create software solutions presently but while analytics might be handling the problem on the surface, this model is not going to last with the vast amounts of data soon to be heading our way.

Mobile is not a trend. What is important about mobile is how consumers are using their mobile phones not just for personal utility, but also for commercial reasons. Businesses are starting to plough money into it as a platform, understanding that there is ample

opportunity in this ever-evolving channel. To remain competitive, there is a need to focus on mobile analytics, allowing for improved user experience.

### **Read: The Big Data Collection Problem of Little Mobile Devices.**

Slightly off-track but incredibly interesting if you are into the legality of data collection. This looks at mobile and is fascinating to see how information can legally and commercially be collected, and uncovers a few ways in which applications are used on mobile devices that you may not be aware of!

[Download here. \(Links to an external site.\)](#)

### **Fraud Detection and Security**

Fraud is intentional deception made for personal gain or to damage another individual. Anyone who has had their credit-card “swiped” / “skimmed” or details stolen would appreciate the role that Big Data has played in the banking industry in terms of security. Real-time business analytics are assisting here in monitoring credit card transactions, monitoring and checking in near-real time. If the checks identify pattern inconsistencies and suspicious activity, the transaction is identified for review and escalation, often with sophisticated (and fast!) methods of processing.

Big Data technologies provide an optimal technology solution due to the reliance on large dataset figures, as well as cross-referencing of information. The precision that is required to detect (often globally in terms of real-time data collection and processing) is also incredibly important in this age of integrated markets and e-commerce. Another emergent approach to solving fraud with Big Data is social network analysis (SNA). SNA is the precise analysis of social networks which view social relationships and make assumptions. SNA can effectively reveal all individuals involved in fraudulent activity, from perpetrators to their associates, and attempt to understand their relationships and behaviours to identify an in-progress fraud case.

Javelin’s “2012 Identity Fraud Report: Social Media and Mobile Forming the New Fraud Frontier” found that:

- 68% of people with public social media profiles shared their birthday information (with 45% sharing month, date, and year)
- 63% shared their high-school name; 18% shared their phone number
- 12% shared their pet’s name — all are prime examples of personal information a company would use to verify your identity.

It is not uncommon to hear top analytics professionals state that they work in risk management (RM). Almost unequivocally, risk management is now data-driven. The two most common types of risk management are credit risk management and market risk management. A third type of risk, operational risk management, is not as common, but is becoming increasingly visible in the RM space. The tactics for risk professionals

typically include avoiding risk, reducing the negative effect or probability of risk, or accepting some or all of the potential consequences in exchange for a potential upside gain. The vast amount of both qualitative and quantitative information available to credit risk professionals can be overwhelming to digest and can slow down a process with potential sales at risk. With advanced analytical tools, these abundant and complex data sources can be distilled into simple solutions that provide actionable insights and are relatively easy to implement.

Credit risk analytics focus on past credit behaviours to predict the likelihood that a borrower will default on any type of debt by failing to make payments which they obligated to make. For example, “Is this person likely to default on their \$300,000 mortgage?” Market risk analytics focus on understanding the likelihood that the value of a portfolio will decrease due to a change in stock prices, interest rates, foreign exchange rates, and commodity prices. For example, “Should we sell this holding if the price drops another 10 percent?”

## Commerce and Trading

There has been a significant shift in the use of data in the financial services industry over the past decade, with financial institutions at the forefront of applying these analytics for a variety of commercial and trading functions: from risk management, proprietary trading to portfolio management. Investment banks, in particular, are adept at using highly sophisticated algorithmic trading systems. These processes are enabled by machines built on complex algorithms and mathematics to determine buy and sell orders for equities, commodities, interest rate and foreign exchange rates, derivatives, and fixed-income instruments at as close to instant as possible.

A key component of algorithmic trading is determining return and the risk of each potential trade, and then making a decision to buy or sell. Quantitative risk analysts help banks develop trading rules and implement these rules using modern technology. Speed is obviously a priority in an era of digitised commodity and stock trading, and increasingly, decisions are “automated” via systems of intense calculation.

The multiple applications we have canvassed here also require and demand multiple skills in the decision-making team, multiple tools, and multiple types of data (clickstream data, consumer data, competitive intelligence data, etc.). Historically, businesses have failed to embrace this variety of data, instead focusing on “good” data, that is clean, manageable and complete

---

## Chapter Navigation

Big Data in Business

[2.0 Chapter Overview](#)

[2.1 Business and Data](#)

[2.2 Data Applications](#)

[2.3 Analytics and Big Data](#)

[2.4 Big Data Evolution and Business Trends](#)

[2.5 Chapter Summary](#)

## 2.4. Big Data Evolution and Business Trends

### Application Evolution

|                                | <b>Improve Operational Efficiencies</b>                                                                                         | <b>Increase Revenues</b>                                                                                   | <b>Achieve Different</b>                                                                                                    |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Mature Analytic Applications   | <ul style="list-style-type: none"><li>• Supply chain optimisation.</li><li>• Marketing campaign optimisation.</li></ul>         | <ul style="list-style-type: none"><li>• Algorithmic trading.</li></ul>                                     | <ul style="list-style-type: none"><li>• In-ho<br/>applic</li></ul>                                                          |
| Maturing Analytic Applications | <ul style="list-style-type: none"><li>• Portfolio optimisation.</li><li>• Risk management.</li><li>• Next best offer.</li></ul> | <ul style="list-style-type: none"><li>• Ad targeting optimisation.</li><li>• Yield optimisation.</li></ul> | <ul style="list-style-type: none"><li>• In-ho<br/>applic</li></ul>                                                          |
| Emerging Analytic Applications | <ul style="list-style-type: none"><li>• Chronic disease prediction and prevention.</li></ul>                                    | <ul style="list-style-type: none"><li>• Custom churn prevention.</li></ul>                                 | <ul style="list-style-type: none"><li>• Produ<br/>optim</li><li>• Desig<br/>optim</li><li>• Branc</li><li>• Produ</li></ul> |

These emergent applications of Big Data show the opposite can be true: dynamic, real-time data that is large in set and complex in its format is sometimes the richest in terms of insights available.

It is imperative that if you want to make sound decisions with increasingly digital and online information, organisations must learn the many different types of tools to bring this different type of data together. The table demonstrates the multiple business functions and outcomes versus the evolution of data applications we have canvassed in this chapter. We can see here what existing and what emergent analytic capabilities can provide, not only for businesses but for society.

### Business Trends

## Data Trends

### Governance and Self-Service Analytics become best pals!

Businesses have realised that data governance, when done right, can help nurture a culture of analytics and meet the needs of the business. People are likely to dig into their data, when they have centralised, clean and fast data sources and when they are confident about an IT person looking after security and performance.

### Watch: Citrix Story

Citrix Story from Tableau, a case on how to scale the culture of analytics in business.

### Data becomes democratised

Self-Service Analytics tools have changed expectations of people for good. As millennials enter the workforce, employees will seek empowerment. Data in businesses have to be shaped on the go. Therefore, the demand for self-service data preparation tools and self-service warehouses will grow. Democratisation of data allows people to respond quickly to shifting priorities in business!

### Data Integration is Exciting

Today, business want to connect the right data to the right people, quickly. Data sources are varied and there is an increasing rise of agile tools that enable data explorers to connect to any dataset, wherever it lives. It is predicted that the growth of agile tools will increase over the next 10 years.

### Data is no longer an analyst's domain

All business people are expected to become data savvy, to obtain meaningful experiences. Platforms that enable users to apply statistics, ask a series of questions and stay in the flow of analysis, will become mainstream.

### Cloud takes off!

Cloud storage and retrieval has now become commonplace. Cloud analytics helps businesses become agile and they have become more reliant on it – similar to any other enterprise systems. (Chapter 2 will cover more on Cloud platforms).

### COE becomes excellent!

Many business are establishing Centre of Excellence (COE) to foster adoption of self-service analytics. There are also enablement programs such as online training forums and one-to-one training. These centers enable data to inform workflows across the organisation in the long term.

## Mobile analytics gains recognition

Working with mobile devices and data is no longer an extension of computing. As part of dynamic analytics processes, the world is becoming mobile. Data visualisation using iPads is an example.

## Internet of Things (IoT)

Everything today is poised to have sensors that send information back and forth. Location tracking using sensors is now becoming mainstream and, as it grows, companies will look for tools that allow users to explore data and share findings, securely, but in an interactive manner.

---

### Chapter Navigation

**Big Data in Business**

[2.0 Chapter Overview](#)

[2.1 Business and Data](#)

[2.2 Data Applications](#)

[2.3 Analytics and Big Data](#)

[2.4 Big Data Evolution and Business Trends](#)

[2.5 Chapter Summary](#)

## 2.5. Chapter Summary

What society is doing with Big Data – our thinking about the technology-impact cycle; the order by which technology passes through society and perhaps into the industry. We also noted consumers' own willingness to harness and create data and touched on the quantified self-revolution which presents itself as a pertinent example of the information age and societal adoption.

### Big Data applications

An overview of the various types of data sources, a history of application from ERP to CRM and now into the age of analytics. We also touched upon many industries and the various ways in which Big Data is used, from medicine to advertising and marketing.

## Emerging analytic applications

The emerging Big Data universe was explored with the many possibilities around mobile, technological innovation and online web 3.0 development. We took a deeper dive into the mobile/web and network research areas around analytics usage, and some of the major impacts on the commercial, trading and finance sector of these new innovations around informatics and analytics. The evolution and maturity over time of business analytics applications were used to conclude this exploration of contemporary methods.

## Trends

Here we touched upon some trends such as mobile analytics, IoT, data democratisation, integration and the creation of Centres of Excellence for Data. We explore these much more in detail in Chapters 2 and 3.

## Exercise

Try to answer these review questions briefly:

1. Having looked at the evolution of business analytics and its various applications, what do you think are the three most important functions of these applications for industry? Why?
2. Why is there so much conjecture around the conceptualisations of Web 2.0 and Web 3.0. What are the key differences? Moreover, what new opportunities might be presented to new businesses or enterprises with these advances?
3. What do you think are some of the operational (and perhaps commercial) reasons behind organisations not employing complete marketing automation? What are some of the key benefits?
4. Explain what you think the relationship between CRM and Big Data looks like for the modern-day organisation. Do you think these processes and systems are essential?
5. Explain why mobile should not be considered a “trend” for industry. Think closely about the current “information era” and what has changed in the way business is done on this platform.

Answer here.

## References

1stwebdesigner (n.d.). A brief introduction to Web 3.0,  
<http://www.1stwebdesigner.com/web-3-introduction/>

Investopedia (2015) <http://www.investopedia.com/terms/q/quantified-self.asp#ixzz3yIFK3s8H>

LinkedIn Learning (2016) Data Science and Analytics Career Paths and Certifications

Marr, B. (2015). Big Data: Using SMART Big Data. Analytics and Metrics to Make Better Decisions and Improve Performance.

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Redman, T. C. (2008). Data driven: profiting from your most important business asset. Harvard Business Press.

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Tableau (2016) Top 10 trends for 2016 - Business Intelligence

Van Rijmenam, M. (2014). Think Bigger: Developing a Successful Big Data Strategy for Your Business. AMACOM Div American Mgmt Assn.

## Extra Resources

| Resource                                                                                                                                                                                  | Description                                                                                                                                                                                                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Light, easy and fun                                                                                                                                                                       |                                                                                                                                                                                                                                                      |
| Barns (2015) "Big Data's Big Impact on the Future of Advertising" in re/code                                                                                                              | Barns, who is the CEO of Nielsen (the la and insights firm), gives a succinct defec about 'big data', but the 'right data'.<br><a href="#">Read here (Links to an external site.)</a>                                                                |
| Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. McKinsey Quarterly, 4, 24-35.                                                                         | A slightly outdated (2011) article on Big interesting predictions. Read this in it's h think about what had changed or not go has a nice graphic on the various indust prospective impact.<br><a href="#">Read here (Links to an external site.)</a> |
| For the academically curious                                                                                                                                                              |                                                                                                                                                                                                                                                      |
| Issa, N. T., Byers, S. W., & Dakshanamurthy, S. (2014). Big data: the next frontier for innovation in therapeutics and healthcare. Expert review of clinical pharmacology, 7(3), 293-298. | A top-line and straightforward overview Data and informatics on pharmacology. short, overview of the enabling potential industry.<br><a href="#">Download here (Links to an external site)</a>                                                       |
| Arnold, M., & Kiker, D. R. (2015). The Big Data Collection Problem of Little Mobile Devices. Richmond Journal of Law & Technology, 21(3).                                                 | Slightly off-track but incredibly interesting legality of data collection. This looks at r fascinating to see how information can l commercially be collected, and uncover                                                                           |

| Resource | Description                                                                                                                                                                        |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|          | <p>which applications are used on mobile devices and how to use them, and how to use them to not be aware of!</p> <p><a href="#">Download here</a> (Links to an external site)</p> |

---

## Chapter Navigation

[Big Data in Business](#)

[2.0 Chapter Overview](#)

[2.1 Business and Data](#)

[2.2 Data Applications](#)

[2.3 Analytics and Big Data](#)

[2.4 Big Data Evolution and Business Trends](#)

[2.5 Chapter Summary](#)

---

## Chapter 3

### 3.0. Chapter Overview - Big Data Value Creation

In this chapter, we look at the dimensions of Big Data and the process of value creation. The Big Data maturity model is considered and the complete value creation model is presented.

1. Dimensions of Big Data.
  2. Value Creation Process.
  3. Theoretical Propositions.
  4. Value Creation Model.
- 

## Chapter Navigation

## Big Data Value Creation

### [3.0 Chapter Overview](#)

### [3.1 Dimensions of Big Data](#)

### [3.2 Value Creation Process](#)

### [3.3 Theoretical Propositions](#)

### [3.4 Value Creation Model](#)

### [3.5 Chapter Summary](#)

## 3.1. Dimensions of Big Data

### Image: Dimensions of Big Data

#### 2016 Gartner Hype Cycle

In Gartner's model, "Big Data" as a catch-all term disappears, but is instead fragmented into various categories including "advanced analytics", hybrid cloud, microdata centres and citizen data science. Some of these have been discussed in previous chapters, while others are coming up. What is important here is the shift in terms of how the industry has evolved: the strategic value of data and analytics is being increasingly recognised, and how quickly and readily industry has adopted aspects of Big Data and business analytics into everyday processes.

We previously discussed the definitions and conceptualisation of Big Data - what it means and how it is described (from a variety of perspectives).

Remember the 7Vs of Big Data - Volume, Velocity and Variety (and) Rijmenam's Veracity, Variability, Visualisation, and Value (2014). This helped us better conceptualise and understand the impact and implications of Big Data on business strategy. These 7Vs however are not perfect at encapsulating the technical complexity that stems from the fact that complex predictive and prescriptive analytic methods need to be applied to huge, heterogeneous data sets. To extract the most value from these data sets, the challenges and opportunities of Big Data can be viewed across five distinct dimensions: technology, application, economic, legal and social.

### Image: The Five Dimensions of Big Data

| Dimension  | Description                                                                                                                                                                         |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Technology | There is a need for scalable systems and platforms for data analysis (for datasets of imm as insightful and original data analysis methods To help crunch these effectively. Techno |

| Dimension   | Description                                                                                                                                                                                                                                                                                                                                                               |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|             | needed to assist in overcoming a distinct skills gap to enable data analysis methods to be used by wider audiences.                                                                                                                                                                                                                                                       |
| Application | Many emergent applications are arising in the information economy, such as information being refined and sold enriched data. These information marketplaces are effectively building the information economy from scratch. Other examples were discussed in the last chapter, such as the quantified self, medical and healthcare breakthroughs and Web 3.0 applications. |
| Economic    | Value creation in the economic dimension lies in new business models and enterprises, as content is delivered, bought, and sold. For example, think about paradigm shifts around the source software, as well as the marketplace for information as commodity (especially in the financial environments such as trading).                                                 |
| Legal       | From a legal perspective, Big Data will present many challenges with respect to ownership, insolvency, in addition to prevalent issues around privacy and security. Some of these will be discussed in the course.                                                                                                                                                        |
| Social      | Lastly, Big Data and analytics will have a profound impact on society as a whole with respect to interaction, news, and democratic processes, among others. The creation of societal-value cannot be ignored, especially in the role of social enterprise.                                                                                                                |

### Image: Creating Smarter Cities

The following short Forbes article details ways in which technological advancement has led to a proliferation of data and how to make our cities operate more efficiently.

[Read here. \(Links to an external site.\)](#)

1. Explain and dot point each of the five dimensions you think are applicable to this article. (Think closely about value creation from an economic and social perspective.)
2. What is the value of real-time analytics and monitoring here to both private and public interests? Explain why you think investment is so forthcoming.
3. Search for similar applications of the Internet of Things in an industry that interests you. How is value being created from this technological advancement?

[Answer here.](#)

## Chapter Navigation

[Big Data Value Creation](#)

[3.0 Chapter Overview](#)

[3.1 Dimensions of Big Data](#)

[3.2 Value Creation Process](#)

[3.3 Theoretical Propositions](#)

[3.4 Value Creation Model](#)

[3.5 Chapter Summary](#)

## 3.2. Value Creation Process

### The Five Dimensions of Big Data

As innovators and business decision makers, it is important to understand how value is - and can be - created across and between these different dimensions. Indeed, data management and value creation for technological advancement may look different from the ways in which benefits are perceived from a societal perspective, and might come into conflict with other dimensions (such as legal or economic).

One of the key challenges organisations face in building support for a Big Data initiative is to ensure that the Big Data initiative is valued by, or of value to, as many stakeholders as possible. And, rather unfortunately, stakeholders have become increasingly jaded with promises of the next great technological advancement in terms of business and market growth potential. They are hesitant to believe that another new technology is going to solve all of their data and analytic problems. You are now going to be introduced to a method that will ensure the business relevance of your Big Data initiative to maximise value for involved stakeholders.

### What Types of Value? To Whom?

Both in businesses and in science, data use is handled in a fragmented way. Actors along the Big Data Value chain should cooperate and form the basis of a strong and vibrant data-driven ecosystem to maximise value creation of Big Data. The following value chain extends this notion of benefits across all stakeholders, which will only succeed when the individual links operate with needed and available capabilities.

### Value Creation

|                                 |                                                                                                                                                                                              |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Generation and Acquisition | <ul style="list-style-type: none"><li>• Device and sensor networks</li><li>• Structured and unstructured data</li><li>• Event recording</li><li>• Multi-modality</li><li>• Streams</li></ul> |
| Data Analysis and Processing    | <ul style="list-style-type: none"><li>• Pre-processing</li></ul>                                                                                                                             |

|                                                 |                                                                                                                                                                                                                                 |
|-------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                 | <ul style="list-style-type: none"> <li>• In-memory processing</li> <li>• Sentiment analysis</li> <li>• Correlations</li> <li>• Pattern recognition</li> <li>• Real-time analytics</li> <li>• Machine learning</li> </ul>        |
| Data Storage                                    | <ul style="list-style-type: none"> <li>• In-memory storage</li> <li>• Data augmentation and scaling</li> <li>• Validation and elimination of irrelevant data</li> <li>• Consistency</li> <li>• Revision and updating</li> </ul> |
| Data Visualisation                              | <ul style="list-style-type: none"> <li>• Decision support and reinforcement</li> <li>• Modelling</li> <li>• Simulation</li> <li>• Prediction</li> <li>• Exploration and control</li> </ul>                                      |
| Social, Economic, Legal, Technological Benefits | <ul style="list-style-type: none"> <li>• Security</li> <li>• Data protection</li> <li>• Privacy</li> <li>• Trust</li> </ul>                                                                                                     |

Cooperation is needed across all aspects of the data and information management process. All links are bound by overarching security, privacy and trust concerns. The benefits here can be across all dimensions of the Big Data framework as discussed in the previous sections. This can be the beginning of a healthy and vibrant approach to value creation from your data, with the maximum benefits to most, if not all, actors and stakeholders involved.

### Read: Creating Value Through Advanced Analytics

Bain & Co. always have a unique spin on things, and this page and report is very good at stressing the business case for value creation.

[Read here.](#) (Links to an external site.)

### Image: Value Creation Process

### Big Data Business Model Maturity Index

When we talk about value creation, we also talk about business enablement and efficiency.

- How far can Big Data take us from a business perspective? What is the end goal?
- How far can I push Big Data to power, or even transform, my value-creation processes?

The previous value chain demonstrated the benefits of businesses moving from a retrospective view of the business with partial chunks of data in batch to monitor their business performance, to an environment that integrates predictive analytics with real-time data feeds that leverage all available data in order to optimise the business function. The following Big Data Business Model Maturity Index, developed by Schmarzo (2013) is useful in helping us identify where organisations currently sit, and map out where they could be (their desired state) with respect to leveraging Big Data to uncover new monetisation and business development opportunities.

The first three phases of the Index are internally focused — optimising an organisation's internal business processes. This part of the maturity index leverages an organisation's data warehouse and business intelligence investments, especially the key performance indicators and data models.

The last two phases of the Index are externally focused — creating new monetisation opportunities based upon the customer, product, and market insights gleaned from the first three phases of the maturity index. This is the part of the Big Data journey that catches most organisations' attention: the opportunity to leverage the insights gathered through the coordination of their internal business processes to create new business and value opportunities.

(Schmarzo, 2013)

### Watch: Global Leaders Faculty

Mr. Rich Caccappolo, COO of MailOnline, explores the challenges in creating value from data.

### Watch: Amazon

Watch this short video on this history of Amazon.

[Watch here.](#) (Links to an external site.)

Then, read about Amazon's latest efforts in North America.

[Read here](#) (Links to an external site.)

1. Think closely about the Big Data Business Model Maturity Index, and Amazon's current position in terms of integration of analytics and monetisation of this data. Where would you currently place them on the index? Why?
2. Afterwards, perhaps plot Amazon's trajectory over the past 20 years in terms of internal and external focus and how this has led to successful value creation. Consider what is next for Amazon. How do they capitalise on this information? Does it require internal optimisation, or capitalisation on external market insights?

[Answer here.](#)

---

## Chapter Navigation

### Big Data Value Creation

#### [3.0 Chapter Overview](#)

#### [3.1 Dimensions of Big Data](#)

#### [3.2 Value Creation Process](#)

#### [3.3 Theoretical Propositions](#)

#### [3.4 Value Creation Model](#)

#### [3.5 Chapter Summary](#)

## 3.3. Theoretical Propositions

### Image: Theoretical Propositions

## How Can Data Be Turned Into Value?

The previous Business Model Maturity Index explains one such pathway, but what are the components of these stages? What does “business optimisation” mean when we are dealing with analytics, data and information? The first and most common proposition for successful value creation is acceptance of the following: An organisation's business analytics capability mediates between its data and the creation of value. (see above image). Here, organisations extract and promote value (across many dimensions, see previous pages) because of their capabilities and some of their systems in place.

More specifically, with regards to what organisations should do if they want to create value from their data, the model above proposes an analytics “eco-system” that places the business analytics function in a much broader organisational context. Three key dimensions here, which we will explore further, are data, management and technology.

This theoretical position helps us explore the following pertinent questions:

1. What dimensions do organisations need to address when building a business analytics capability?
2. What are the organisational change implications of building and exploiting a business analytics capability?

## Data

What dimensions do organisations need to address in building a business analytics capability?

| Dimension                                    | Key Consideration                                                                                                                                                          |
|----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ensure good quality data                     | All data is ready and fit for task purpose.                                                                                                                                |
| Appropriate permissions and security         | Transparency is needed not just for accountability, but for customer trust. Trust is maintained and confidence established through these safeguards with all stakeholders. |
| Sharing mindset                              | All outcomes, both positive benefits and negative externalities (i.e. data extraction and analytical process) are shared across all appropriate stakeholders.              |
| Building data partnerships                   | Value comes from this shared approach: partnerships because of shared data rather than viewing of data as a saleable commodity.                                            |
| Creation of both public and private benefits | Data is used for both public and societal benefit, as well as commercial operations optimisation, versus public weather alerts and warnings.                               |
| Appropriate regulation and control           | Changes in legislation may result in fundamental shifts in what can be done with customer data (consumer rights to collection/contact etc).                                |

## Management

What dimensions do organisations need to address in building a business analytics capability?

| Dimension                           | Key Consideration                                                                                                                                  |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Robust corporate analytics strategy | A strategy that is clear and articulated about where value will be created and how.                                                                |
| Embracing organisational change     | Agility across both hardware and software organisational units, underpinned by a data-driven cohesive cultural shift.                              |
| Dynamic sector knowledge            | Deep understanding of the organisation and its place in the industry (vs. competitors) in terms of analytics advancement and technological change. |
| Organisational structure            | Teams and units are cross-functional and contain a mix of data science and business expertise.                                                     |

| Dimension                    | Key Consideration                                                                                                        |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------|
|                              | thinkers and IT specialists. Teams should have resources available for analytics opportunities.                          |
| Innovative Partnerships      | Proactive partnerships with key think-tanks and institutions, including faculties in universities and data laboratories. |
| Ethics and process diligence | Review committees should be consistently convened to provide data is used to protect the reputation of organisations.    |

## Technology

What dimensions do organisations need to address in building a business analytics capability?

| Dimension                              | Key Consideration                                                                                                                                                                                              |
|----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Visualisation Tactics Across Functions | Using visualisation to assist in explaining key business plans and making the data accessible for all business units.                                                                                          |
| Embracing Technological Dynamism       | Understanding the dynamic pace of technological change and er from over-capitalisation in risky IT investments, while promoting innovation and excitement in advances.                                         |
| Capable and Curious Data Scientists    | The people working with your data must be able to work independently and sets and tools to manage these, as well as able to process this for actionable insights.                                              |
| Assemblage Mindset                     | Teams and units are able to find solutions using a variety of analysis and interpretation methods. Organisations must encourage and facilitate types of tools to bring these different types of data together. |
| Acquisition and Retention of Key Staff | Data analytics teams are retained and managed closely and have progression and paths of advancement to encourage personnel commitment to current projects.                                                     |

**Read: What's the social value of the arts? Big Data has some answers (even if artists don't want to hear it)**

I will balance this out with this nifty short piece by Taylor which is arguing the opposite to the above...how can we measure the social benefit or value of sectors like the Arts. An interesting footnote in how data and analytics may help answer these questions.

[Read here.](#) (Links to an external site.)

## Case Study: City Trans

This case study details not only a sophisticated integration of Big Data and analytics processes, but also a unique emphasis on end-user value creation.

CityTrans is a governmental, not-for-profit provider of an integrated public transport system for a major city, dealing with every aspect of how people move across the city using different modes of public and private transport. The organisation works with many datasets, such as network operations, travel data, traffic data, loadweigh data, infra-red data, and customer data. Some of these data assets have been linked for operational analysis and planning but there are still some unexplored opportunities in joining these numerous and diverse datasets.

CityTrans collects the bulk of its public transport travel data through a smart travel card (STC), which can be used anonymously or as a registered (and therefore identifiable) customer. Any one passenger may have a number of anonymous, pay as you go or low-value season ticket STCs, making identification of an individual difficult. To further protect individuals' identities CityTrans' data scientists and research partners use pseudonymised ("hashed") IDs for their analyses of customer behaviour.

The STC generates millions of "taps" a day as customers use it on different modes of transport throughout the city. This data typically tells CityTrans where a passenger entered the system and where they left it but it is more difficult to determine what route they took through the network. CityTrans also accepts contactless payments from credit and debit cards on its bus network and has plans to roll out contactless payment to the rest of the public transport network. CityTrans has been active in data analysis and modelling of passenger flow for many years, with analysis grounded in traditional operational research methods. Making use of data has become more prominent since the introduction of the STC and as storage capacity has increased and new software tools have become available.

## Customer Value Creation

The data collected from users of the transport system and the system operations are used to measure the reliability of the service and to gain insight into the customer experience. Rather than focus on average journey times as a proxy for customer satisfaction, CityTrans is now able to use the STC data to get closer to the customer experience and to use the data to minimise travel delays.

When there is disruption to the network, a measure of central tendency, such as the average or the median, is a poor representation of the impact on passengers. CityTrans is also investigating the use of data for behavioural change, to shift travel patterns so as to spread the load across the network. CityTrans envisages changing traveller behaviour using "recommender" systems that will assist the traveller in planning their journeys. (Vidgen, 2014)

### Partnerships

In order to enable even better use of STC data, CityTrans has established research partnerships with several universities, to explore their data and apply new analytic techniques, such as visualisation. A relationship with MIT

### Processes

The decision to develop a new data war CityTrans to look at its processes for ex transforming and loading data, and in re align them with best practice for busines

## Partnerships

has proved particularly effective.

## Processes

systems development. A modified Agile is used for development.

The previous sections illustrate three areas on which businesses must focus when building their analytics and data capability. From a value creation perspective, while all need to be in alignment, we can see key dimensions here that are distinct but intertwined with reciprocal relationships. For example, the technological dimension may constrain what can be done with available data. At the same time, the data held and the analytics strategy will influence the management approach. The analytics eco-system illustrates the mutual co-dependencies that are of particular interest when defining an analytics strategy.

This goes some way to looking at how we enable organisations and businesses to adopt this focus on data systems and analytics, however, the second question we needed to ask is:

What are the organisational-change implications of building and exploiting a business analytics capability?

---

## Chapter Navigation

### Big Data Value Creation

[3.0 Chapter Overview](#)

[3.1 Dimensions of Big Data](#)

[3.2 Value Creation Process](#)

[3.3 Theoretical Propositions](#)

[3.4 Value Creation Model](#)

[3.5 Chapter Summary](#)

## 3.4. Value Creation Model

### Image: Value Creation Model

There are three fundamental questions businesses have for decades had to answer:

- Who are the most valuable and important customers?
- How do I reach them?
- With which products?

Schmarzo's illustration highlights how Big Data drives value creation. Big data analytics has changed how these new sources of data can be leveraged to answer these key business questions. For example, we can better direct R&D thanks to shifting our focus from the most "profitable" customers to the most "influential". Big Data enables you to answer these questions not only with a higher level of detail but at faster speeds, greater precision, and further insight.

## A Complete Model of VC

The transition to a data-driven organisation in which business decisions are routinely made on the basis of data and hypothesis testing will require deep changes to processes, management, structure, and ultimately to the culture of the organisation. Verhoef et al's model here of value creation nicely encapsulates these various dimensions and influencing factors.

| Big Data Assets                                          | Big Data Capabilities                                                                                                       | Big Data Analytics                                                                                                                                                                      | Big Data                                                                   |
|----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Data</li> </ul> | <ul style="list-style-type: none"> <li>• People</li> <li>• Systems</li> <li>• Processes</li> <li>• Organisations</li> </ul> | <ul style="list-style-type: none"> <li>• Insights</li> <li>• Models</li> <li>• Decision/Support</li> <li>• Actions/Campaigns</li> <li>• Information Based Products/Solutions</li> </ul> | <ul style="list-style-type: none"> <li>• Value</li> <li>• Value</li> </ul> |

### Watch: Global Leaders Faculty

Mr John Fraser, Secretary of the Department of Treasury, Australia discusses how we should value past and projected performance in present terms.

## Not To Be Used in Isolation

Throughout your studies you have been exposed to many other models of Value Creation, notably, Michael Porter's Valuation Creation Models. These cannot be discounted for the insights they provide especially in regards to resources and capabilities, when looking at process models such as:

- Five Forces Analysis
- Value Chain Analysis

These illustrate impacting factors of the competitor space, as well as some of the internal activities of a business. Think closely here about what these models aim to illuminate, as many of these insights can be combined with the extra benefits and solutions that a comprehensive Big Data and analytics strategy can provide for even

greater outcomes. Value creation should be seen as a process: inherently “value” as a concept is dictated by market exchanges. This process can, and should be, supported and supplemented through rigorous understanding of how these exchanges operate, and interact. Ultimately, the focus here is on how value can be added from these very exchanges, and the impact on the organisation, its customers, and society more widely.

## So What?

For businesses to maximise the benefits of their analytics programs, data and value should be considered and managed jointly. To create value data has to be managed (remember: quality of the data, permissions and regulation, etc.) and that value may have to be shared with those who provide the data.

Secondly, business analytics is not a technical project to be given only to the IT department – it is a total organisation-wide transformation that requires a strategy, management support, and active and careful change management. This is not to say that IT is unimportant but we must recognise the role of data and analysis teams as enablers of the business analytics process and essentially embedded in the organisation’s wider processes and practices.

Thirdly, the organisation needs a healthy respect for change to implement a robust business analytics strategy. This will be expanded later in the course when we look at the organisation and decision making regarding learnings and strategy from Big Data strategy.

Fourth, data scientists need a strong sense of independence and problem-solving ability, as well as the ability to combine tools and techniques towards novel and actionable insights.

And finally, business analytics and data science needs a process model, which acknowledges the organisational context to define and structure these efforts described, to extract (or “capture”) the most value from its activities.

### **Read: Big Data Alchemy: How Can Banks Maximise the Value of their Customer Data.**

A solid report put out by the Capgemini group which looks at the role of Big Data analytics in the banking sector and nicely demonstrates the role data in aiding the customer-bank relationship. What I love here is the roadmap to analytics maturity they include, perhaps consider it a more detailed version of the index we canvassed in this chapter!

[Read here.](#) (Links to an external site.)

---

## Chapter Navigation

### Big Data Value Creation

[3.0 Chapter Overview](#)

[3.1 Dimensions of Big Data](#)

[3.2 Value Creation Process](#)

[3.3 Theoretical Propositions](#)

[3.4 Value Creation Model](#)

[3.5 Chapter Summary](#)

## 3.5. Chapter Summary

In this chapter we looked at the following:

### The Five Dimensions of Big Data

Introduction to the five dimensions of Big Data and the various challenges and opportunities each can present. To extract the most value from data sets, the challenges and opportunities of Big Data can be viewed across technology, application, economic, legal and social dimensions.

### The Big Data Business Model maturity index

We looked at this index to get a sense of growth and trajectory of analytics incorporation. It is useful in helping us identify where our organisations currently sit, and map out where they could be (their desired state) with respect to leveraging big data to uncover new monetisation and business development opportunities.

### Models of Value Creation

Here we adopt a theoretical perspective, exploring some of the models of data system incorporation, and discuss the methods by which these insights are processed by businesses.

### Coordination of Big Data for VC

We finish the chapter with a complete process model, taking into account the above chapter components. This roadmap will provide you with a template for effective implementation of data and business analytics plans, and can be used in your assignments, and for your own entrepreneurial ideas and planning. It is important that as business leaders we build an analytics strategy that clearly articulates data sources, emphasises data quality, and enables cultural change. Analytics projects do not simply

concern themselves with IT departments. Entrepreneurs should learn and incorporate these Big Data and analytics processes quickly, cheaply, and with agility in response to technological and market shifts.

## References

Gartner (2015) "Gartner's 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor", accessed online at: <http://www.gartner.com/newsroom/id/3114217> [January, 2016]

GE's Global Innovation Barometer, 2016 , accessed online at: <http://www.forbes.com/sites/louiscolumbus/2016/01/20/ge-global-innovation-barometer-2016-61-of-execs-using-big-data-analytics-to-improve-decisions/#4b4233f35f89> [January, 2016]

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Van Rijmenam, M. (2014). Think Bigger: Developing a Successful Big Data Strategy for Your Business. AMACOM Div American Mgmt Assn.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge.

Vidgen, R., (2014). Creating business value from Big Data and business analytics: organisational, managerial and human resource implications. Hull University Business School Research Memorandum, no. 94

## Readings

| Resource                                                                                                               | Description                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Light, easy and fun                                                                                                    |                                                                                                                                                                                                                                                                                                             |
| Bain (2015) "Creating Value Through Advanced Analytics"                                                                | Bain & Co. always have a unique spin on their page and report is very good at stressing the importance of analytics for value creation. I dare you to read this report on analytics the primary tool for strategic advantage. To pitch, this one.<br><a href="#">Read here (Links to an external site.)</a> |
| Taylor (2014) "What's the social value of the arts? Big Data has some answers (even if artists don't want to hear it)" | The article by Taylor presents an alternative view on how Co...how can we measure the social value of sectors like the Arts. An interesting footnote                                                                                                                                                        |

| Resource                                                                                                                        | Description                                                                                                                                                                                                                                                                                                                                                                                                             |
|---------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                 | analytics may help answer these questions.<br><a href="#">Read here. (Links to an external site.)</a>                                                                                                                                                                                                                                                                                                                   |
| Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science, 343. | An excellent article looking at the role big data (and Google) can play in solving social ills, though it does not mention the role of social media. This is a well-cited article and nicely explains the dangers of hubris, transparency, as well as expectations of technological determinism.<br><br><a href="#">Read here. (Links to an external site.)</a>                                                         |
| Big Data Alchemy: How Can Banks Maximise the Value of their Customer Data.                                                      | A solid report put out by the Capgemini Consulting. It looks at the role of Big Data analytics in the bank-customer relationship and nicely demonstrates the role data in aiding the bank-customer relationship. What I love here is the maturity of the analytics maturity they include, perhaps the most detailed version of the index we canvasa.<br><br><a href="#">Download here. (Links to an external site.)</a> |

---

## Chapter Navigation

### Big Data Value Creation

[3.0 Chapter Overview](#)

[3.1 Dimensions of Big Data](#)

[3.2 Value Creation Process](#)

[3.3 Theoretical Propositions](#)

[3.4 Value Creation Model](#)

[3.5 Chapter Summary](#)

---

## Chapter 4

### 4.0. Chapter Overview - Information Systems

This chapter demonstrates the pervasive power of ICT and mobile applications, along with cloud storage integration that renders it necessary to examine large population samples in real time. Internet of Things (IoT) and mobile computing are closely examined.

1. Computing.
2. Data and IS.
3. Mobile IS.
4. IoT.

## Quote

“You can have data without information, but you cannot have information without data. ”

Daniel Moran, programmer and author

---

## Chapter Navigation

### Information Systems

#### [4.0 Chapter Overview](#)

#### [4.1 Computing](#)

#### [4.2 Data and Information Systems \(IS\)](#)

#### [4.3 Mobile Informations Systems \(IS\)](#)

#### [4.4 Internet of Things \(IoT\)](#)

#### [4.5 Chapter Summary](#)

## 4.1. Computing

### Foundations of Computing

The invention and subsequent proliferation of computers have said to have the single biggest impact in terms of invention since the harnessing of electricity. Almost every aspect of our lives are now mediated in some ways by computers (and, by extension, ICTs). However, we cannot forget the age-old mantra, “Computers are no smarter than the humans who program them”. This is especially pertinent for this course focus: on the role of using this data and learning from insights.

The four main components of any computing device most typically include:

| Component          | Description                                                                                                                                                                                                                                                          |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input Devices      | Input devices allow us to enter information into the computer system from a plethora of sources such as: keyboards, sensors, microphones, etc.                                                                                                                       |
| Processing Devices | Processing components manipulate the information once it has been entered into the computer. There are usually a set of common components consisting of a central processing unit (CPU), interface components, and memory (RAM).                                     |
| Storage Devices    | Storage devices store all of this entered information and various other data for future use. Common storage devices include hard disks and drives, CD-ROMs, and flash storage devices, as well as now cloud based servers.                                           |
| Output Device      | Output devices are how the manipulated information is returned to the user from the computer. They commonly include screens and monitors, printers, and other physical peripherals. They can also include sensors and haptic devices, such as vibrations and sounds. |

Watch: Global Leaders Faculty

Mr Dominic Barton, Managing Director of McKinsey & Co, discusses the exponential growth of computing power and data.

## Measuring Data

Data is measured by basic units of measure that work up from a bit. A bit is represented by either a 1 (electricity flowing) or a 0 (no electricity flowing). This is called binary code. The code converts images, text, and sounds into numbers in order to send information from one digital device such as a computer to another. Computers use binary numbers because they are easier to handle. In binary, the digits (read and write) are worth 1, 2, 4, 8, and so on—not units, tens, and hundreds. A byte is a unit of measure and it is 8 bits put together. In ordinary numbers, “1,001” is one unit. But in binary, “1001” is one 1, no 2, no 4, and one 8, which equals 9.

Binary data is considered the native data/language of a computer and it interacts with the lowest abstraction layer of its hardware. This type of data is produced whenever a process is performed on a computer. The application requesting the process sends instructions in a high-level language that is ultimately converted into binary data to be executed or sent to the processor. All processes, regardless of their type, are converted into binary form before execution. The American Standard Code for Information Interchange (ASCII), uses a 7-bit binary code to represent text and other characters within computers, communications equipment, and other devices. Each letter or symbol is assigned a number from 0 to 127. For example, lowercase "a" is represented by 1100001 as a bit string (which is 97 in decimal).

Watch: Global Leaders Faculty

Mr Ray Sharma, Executive Managing Director of Extreme Venture Partners, explores how big data is a competitive advantage

## Data Processing

Now that we have an understanding of exactly what data is, we need to think about how it is processed; how do computers, and the users that operate them, collect and store data sets? The two types relevant to our understanding of analytics: Batch and Real Time (RT) Processing.

| Batch Processing                                                                                                                                                                                                                                                                                                                                                                                                                        | Real-Time Processing                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>• The collection and storage of data for processing at a scheduled time, when a sufficient amount of data has been accumulated.</li><li>• Data is collected and processed together (multiple transactions)</li><li>• This can be done without human intervention</li><li>• However, there may be a long time delay</li><li>• For example: Generation of bills, credit card transactions</li></ul> | <ul style="list-style-type: none"><li>• The immediate processing of data as it occurs, with the database being updated as the event.</li><li>• Data is collected and processed instantly</li><li>• The act of data processing here can be instantaneous</li><li>• Because of the immediacy, can stop events from occurring</li><li>• Always up-to-date systems</li><li>• For example: Guidance systems in transport, sales and reservations, POS (point of sale)</li></ul> |

### Watch: How We Teach Computers to Understand Pictures

Think about this closely in terms of unstructured data is recorded, processed and stored. Also at the future of computing and data systems that are so visually-based.

## From Data to Systems

When large amounts of data are written and processed together, information is being created. When we talk about information being used functionally (ie: information being computed, for purpose), we talk about systems. A system is a functional unit which involves a set of procedures/functions to produce certain outputs by processing given data and information (as input). More specifically, it is the study of complementary networks of hardware and software that people and organisations use to collect, filter, process, create and distribute data. A computer Information System (IS) is a system composed of people and computers that process or interpret data.

As such, information systems inter-relate with data systems on the one hand and activity systems on the other. An information system is a form of communication system in which data represent and are processed as a form of social memory. An information system can also be considered a semi-formal language which supports human decision making and action – extremely important when considering the applications of data and analytics. Information systems are the primary focus of study for organisational informatics.

## Image: The Traditional IT Systems Approach

Traditionally, in terms of how IT systems are developed and maintained, the requirements are defined, followed by solution design. After this, the system is built, the queries (tasks) are executed. If there are new requirements or queries then the entire system is redesigned and rebuilt.

Consider the system model above.

- Define three key issues with this from an operational or managerial perspective.
- What could go wrong with an IT system designed in this way?
- How does this conflict with our Big Data Value Creation Process model from the last chapter?

| Problem                         | Description                                                                                                                                                                                                                                              |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Limited storage capacity        | We no longer can accept archival data that is historically stored. It is now seen as “data debt” and often is not fit for purpose with today’s systems. It is unable to be integrated and processed due to ineffective methods used in the past.         |
| Limited processing capacity     | Previously we could only process data at limited speeds. Current systems are at processing abilities exponentially beyond what previous systems were.                                                                                                    |
| No scalability                  | Data storage must be scalable. Requirements change and increase over time. Storage must be able to scale itself to new volumes and new data types.                                                                                                       |
| Sequential                      | We cannot process our data in distributed ways on a single machine. Processing is slow. Older systems are sometimes designed around consolidated, batch-focused systems.                                                                                 |
| Data types                      | Older systems can handle structured data well, yet most data generated today is unstructured. We will cover these different types shortly.                                                                                                               |
| Requires pre-processing of data | Previously, data must be stored in RAW form (the most rich form of data). Due to the sheer size of these RAW volumes, older data requires processing before it can be stored efficiently, which sometimes has not been done prior to leaving the source. |

## Image: The Problem With Conventional Computing

Which brings us to the issues we have been summarising thus far. The Big Data problem should now be apparent more so looking at the finer aspects of computing and how data is processed by these systems. This problem is best summarised in the diagram.

---

## Chapter Navigation

### Information Systems

#### [4.0 Chapter Overview](#)

#### [4.1 Computing](#)

#### [4.2 Data and Information Systems \(IS\)](#)

#### [4.3 Mobile Informations Systems \(IS\)](#)

#### [4.4 Internet of Things \(IoT\)](#)

#### [4.5 Chapter Summary](#)

### 4.3. Mobile Information Systems (IS)

#### Image: Mobile

Let us consider the mobile web based information system. Today, we have the ability to access the Internet (and data) via mobile phone, tablets, ipads and a plethora of devices. This requires new forms of connectivity and information exchange – known as Peer-to-peer networks. Proximity based ad hoc exchange (bluetooth enabled), and automatic discovery of peers is possible. For example, transmission of sensor data or traffic data between moving cars.

New requirements and functionality is required such as context awareness and location based services. For example, a restaurant may promote their menu based on your proximity using mobile device location. Geographical Information Systems including maps is another example.

#### Watch: Future of Shopping from Microsoft

---

## Chapter Navigation

### Information Systems

#### [4.0 Chapter Overview](#)

#### [4.1 Computing](#)

## 4.2 Data and Information Systems (IS)

## 4.3 Mobile Informations Systems (IS)

## 4.4 Internet of Things (IoT)

## 4.5 Chapter Summary

### 4.4. Internet of Things (IoT)

Let's start with understanding a few things that are fundamental. Broadband Internet is becoming more widely available, the cost of connecting is decreasing, more devices are being created with Wi-Fi capabilities and sensors built into them, technology costs are going down, and smartphone penetration is sky-rocketing. All of these things are creating a "perfect storm" for the IoT.

#### Image: So What is IoT

This is the concept of basically connecting any device with an on and off switch to the Internet (and/or to each other). This includes everything from cellphones, coffee makers, washing machines, headphones, lamps, wearable devices and almost anything else you can think of. This also applies to components of machines, for example a jet engine of an airplane or the drill of an oil rig. If it has an on and off switch then chances are it can be a part of the IoT. The analyst firm [Gartner \(Links to an external site.\)](#) says that by 2020 there will be over 26 billion connected devices..(some even estimate this number to be much higher, over 100 billion). The IoT is a giant network of connected "things" (which also includes people). The relationship will be between people-people, people-things, and things-things.

And how does it affect businesses? Where are the opportunities? See the image above.

#### Read: OPOWER

Opower is a leading energy management company that combines a cloud-based platform, big data, and behavioral science to help utilities around the world reduce energy consumption and improve their relationship with customers. OPower, in partnership with 93 utilities, helps over 32 million consumer households to lower their energy use and costs and significantly reduce carbon emissions. Working with smart meter, thermostat and other device data from Pacific Gas and Electric, OPower gathers over 7 million data points each day and provides analytic reports to utility companies.

These reports are included in household bills to encourage consumers to conserve energy by comparing their household energy usage to their neighbors. OPower found that their MySQL database infrastructure could not analyze the data quickly enough, and much of the data they were gathering was not being fully used. In fact, while OPower had over 60 instances of MySQL, they still couldn't run analyses across the breadth of their data. To address this issue, OPower created an "Energy Data Hub" by migrating their data infrastructure to distributed systems and using advanced analytics. Their data scientists benefited from self-service, end-user tools that data engineers and product managers could use to access and analyze data in over 200 tables.

Today, OPower product managers use analytics to answer client questions directly – without IT assistance. They can run analyses – for example, using consumer thermostat data to understand patterns of energy usage. The end result? OPower has dramatically lowered the time required to access data for analytics and empowered product managers with insights they used to help clients reduce energy consumption by \$500 million and CO2 output by 7 billion pounds. The case study details details OPower's efforts in giving customers personalised energy usage insights and recommendations.

[Read more here.](#) (Links to an external site.)

Watch: How algorithms shape our world - Kevin Slavin

Another TED this time on the power of algorithms. This one is a bit more esoteric, but still interesting on how computer data systems is infiltrating other aspects of society.

---

## Chapter Navigation

### [4.5 Chapter Summary](#)

#### 4.5. Chapter Summary

| Concept             | Description                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Computation         | We went back to the origins of data and looked at binary code and how it operates in terms of basic code about how all of these 0s and 1s are processed: batch versus real-time and what this means for application and development.                                                                                                                                                  |
| Information Systems | We looked at the evolution of IT systems: from traditional/conventional models which have become unmet challenges of today. We also explored how various systems are designed and operate around data flow and the beginnings of how this can impact on the design of a Big Data strategy                                                                                             |
| Types of Data       | We thought about the key differences between data and information- and how information reveals a different detail, beyond what simplistic data can provide. We also looked at a deeper dive into the variety of data sources of data; both internal and external. We reflected on how an understanding of this leads to different about commercial applications of clever data usage. |
| Mobile IS and IoT   | We explored Mobile information systems and IoT and the opportunities created by them.                                                                                                                                                                                                                                                                                                 |

#### Exercise

Think about the following:

1. What types of data were being collected?
2. What was the source of this data?
3. Do you think there was an emphasis on internal or external data? How were these integrated?
4. Explain the benefits of an IT system design such as this in relation to Big Data.
5. What was the impact on monetisation here; what were the commercial outcomes?

[Answer here.](#)

#### References

Datameer (2015) "Top Five High Impact Use Cases For Big Data Analytics", accessed online at: <http://www.datameer.com/pdf/eBook-Top-Five-High-Impact-UseCases-for-Big-Data-Analytcs.pdf> [January, 2016]

Hadoop Summit (2012) "Combining Hadoop RDBMS for Large-Scale Big Data Analytics", accessed online at: [http://www.slideshare.net/Hadoop\\_Summit/combining-hadoop-rdbms-for-largesacale-big-dataanalytics](http://www.slideshare.net/Hadoop_Summit/combining-hadoop-rdbms-for-largesacale-big-dataanalytics) [January, 2016]

Infogineering (2009) "The Infogineering Model", Image: <http://www.infogineering.net/wp-content/uploads/2009/08/model.jpg> [January, 2016]

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Morgan J (2014) Internet of Things , Forbes. <http://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#64e04f168284>

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Simplilearn (2014) "What is Big Data?", accessed online at: <https://www.youtube.com/watch?v=CKLzDWMSQGM> [January, 2016]

#### Extra Resources

| Resource                                                                                                                                                                                                                                                                        | Description                                                                                                                                                                                                                                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Light, easy and fun                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                              |
| TED Talks, Fei-Fei-Li, How We Teach Computers to Understand Pictures                                                                                                                                                                                                            | Think about this closely in terms of unstructured data stored. Also at the future of computing and data systems based.<br><a href="#">Watch here (Links to an external site.)</a>                                                                                            |
| TED Talks, Kevin Slavin, How Algorithms Shape Our World                                                                                                                                                                                                                         | Another TED this time on the power of algorithms. The esoteric, but still interesting on how computer data systems aspects of society.<br><a href="#">Watch here (Links to an external site.)</a>                                                                            |
| For the academically curious                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                              |
| Ghazal, A., Rabl, T., Hu, M., Raab, F., Poess, M., Crolotte, A., & Jacobsen, H. A. (2013, June). BigBench: towards an industry standard benchmark for big data analytics. In Proceedings of the 2013 ACM SIGMOD international conference on Management of data (pp. 1197-1208). | Ignore the sale-sy pitch in this paper in regards to the instead on how the data was mined/generated, processed into actionable insights here for web visitation and client work together in their model on p1198.<br><a href="#">Read here (Links to an external site.)</a> |

---

## Chapter Navigation

### Information Systems

#### 4.0 Chapter Overview

#### 4.1 Computing

#### 4.2 Data and Information Systems (IS)

#### 4.3 Mobile Informations Systems (IS)

#### 4.4 Internet of Things (IoT)

#### 4.5 Chapter Summary

---

## 5.0. Chapter Overview - Data Warehouse and Cloud

This chapter explores types of data, traditional data stores, distributed systems, database management, and relational databases (RDBMS). Data warehousing methods are then explored, and concepts of OLTP, OLAP are highlighted, indicating the need for real-time processing of data for businesses. The chapter then examines Cloud Computing and solutions/best practices that enable real-time data discovery, that facilitates lean and efficient processes.

1. Storage.
2. Data Warehousing.
3. Cloud Computing.

---

### Chapter Navigation

#### Data Warehouse and Cloud

[5.0 Chapter Overview](#)

[5.1 Storage](#)

[5.2 Data Warehousing](#)

[5.3 Cloud Computing](#)

[5.4 Chapter Summary](#)

## 5.1. Storage

### Finding Homes for Data

We know from the 7Vs of data that Volume is a key component of any Big Data conceptualisation - the sheer amount of data can be seen as overwhelming. Not just for us, but for our systems. Data warehousing techniques have been used for Big Data storage for some time now, however we have seen some developments around the techniques by which we store these large datasets. By far the most common and popular is the relational database (RDBMS) that stores information in rows and columns. In relational databases the primary access is via the row.

Another storage technique is a columnar storage scheme. While a columnar database has rows and columns, information is predominantly accessed via the columns. Lastly,

and most importantly, there are also hybrid approaches, which allow for some information to be easily accessed via row and other data to be easily accessed via columns.

### Watch: Global Leaders Faculty

Ms Angela Toler, Precedent Manager of Arnold Bloch Leibler, explores the importance of efficient storage practices.

### Image: Storage Structures

This illustration demonstrates how this storage is structured.

### Storage Limitations

However it does not come without its limitations - namely, issues of physical space. Some of the most typical and apparent Big Data computing storage limitations encountered are:

| Limitation            | Description                                                                                                                                                                                                                                                                                                       |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Disk bound            | Simply implies there is not enough storage space on the disk. We may be looking at petabytes of data, when only terabytes are available. Side note, here is a nice graphic of the size of Big Data:<br><br><a href="#">View here. (Links to an external site.)</a>                                                |
| I/O bound             | I/O bound refers to the time taken for computation queries to process, indeed “inputs” (I and O, where the name comes from). In terms of a data limitation, this is as a “bandwidth” of sorts, and sometimes there is not enough of this “pipe” to get all the data to meet business needs (in a timely fashion). |
| Memory bound          | Memory is what we use to process these applications and queries. We need to know the data size, and a nice analogy is a blackboard when working out a maths problem. On the blackboard, the more “space” we have to compute our workings.                                                                         |
| CPU bound limitations | Similar in some ways to memory, we use a combination of it and central processing units to process Big Data analytics. Most analytic software today functions by loading data needed for a calculation into memory and then using CPUs to perform calculations.                                                   |

### Overwhelming Systems

The sheer volume and continuous nature of Big Data programs may eventually exhaust your systems if they are not designed accordingly. Appropriate storage is vital for a number of reasons: chiefly, relieving pressure on the human actors in your IT systems. There is a chance that your analytics operators perform shortcuts to process data to compensate for memory or CPU limitations. They may attempt to break apart their

problems to process smaller amounts of the data and instead use this towards achieving their outcomes. They may even use samples and extrapolate. These activities can be costly not only in time (human computation always takes longer) but can also skew your analysis and limit your insights.

Big Data can be a Big Deal - if you are in fact acquiring relevant data and delivering it to specific people within their specific timetables. Call it data merchandising — the act of managing, cleansing and securing highly diverse data from a variety of sources; delivering specific data to specific users; and doing so 24/7 with available personnel and resources. Organisations need a team of IT Analytics Champions that can do this. A team that can efficiently control the data acquisition process and assume responsibility for quality data acquisition and management. The properly positioned IT Analytics Champions don't deliver a barrage of data to everyone at hand; they carefully screen and cleanse that data and then provide pinpoint delivery of that data to exactly the people who need to see it. (Lumidata, 2015)

## Distributed Data Systems

The ability to design, develop, and implement a big data application is directly dependent on an awareness of the architecture of the underlying computing platform, both from a hardware and, more importantly, from a software perspective. Because simplistic traditional computer systems are limited in capacity, they cannot easily accommodate massive amounts of data. That is why high-performance platforms are composed of collections of computers in which the massive amounts of data and requirements for processing can be distributed among a pool of resources. More of these platforms are to come in the next chapter, when we look at Hadoop systems, HDFS and MapReduce.

---

## Chapter Navigation

### Data Warehouse and Cloud

[5.0 Chapter Overview](#)

[5.1 Storage](#)

[5.2 Data Warehousing](#)

[5.3 Cloud Computing](#)

[5.4 Chapter Summary](#)

## 5.2. Data Warehousing

## Image: Data Warehousing

### Current and Historic Data

A data warehouse (DW or DWH), also known as an enterprise data warehouse (EDW), is a central repository of integrated data from one or more disparate sources. They store current and historical data and are used for creating analytical reports for knowledge workers throughout the enterprise. Examples of reports could range from annual and quarterly comparisons and trends to detailed daily sales analysis. The data stored in the warehouse is uploaded from the operational systems (such as marketing or sales). The data may pass through an operational data store for additional operations before it is used in the DW for reporting.

### Data Marts

A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), hence they draw data from a limited number of sources such as sales, finance or marketing. Data marts are often built and controlled by a single department within an organisation. The sources could be internal operational systems, a central data warehouse, or external data.

| Data Warehouse           | Data Mart            |
|--------------------------|----------------------|
| Enterprise-wide data     | Department-wide data |
| Multiple subject areas   | Single subject area  |
| Difficult to build       | Easy to build        |
| Takes more time to build | Less time to build   |
| Larger Memory            | Limited Memory       |

### OLAP/OLTP

Online analytical processing is an approach to answering multi-dimensional analytical (MDA) queries swiftly in computing. It is part of the broader category of business intelligence which also encompasses relational databases, report writing and data mining. Typical applications of OLAP include business reporting for sales, marketing, management reporting, business process management, budgeting, forecasting and financial reporting.

Online transaction processing (OLTP) is a class of information system that facilitates and manages transaction-oriented applications, typically for data entry and retrieval transaction processing. It has also been used to refer to processing in which the system responds immediately to user requests. An automated teller machine (ATM) for a bank is an example of a commercial transaction processing application. Online transaction processing applications are high throughput and insert- or update-intensive in database

management. These applications are used concurrently by hundreds of users. The key goals of OLTP applications are availability, speed, concurrency and recoverability. Reduced paper trails and faster, more accurate forecasts for revenues and expenses are both examples of how OLTP makes things simpler for businesses. However, like many modern online information technology solutions, some systems require offline maintenance, which further affects the cost-benefit analysis of online transaction processing systems.

We Know That:

- Technology has changed.
  - New technologies and open-source inventions enable different approaches that make it easier and more affordable to store, manage, and analyse data.
- The variety of data has evolved.
  - Unstructured data is on the rise, as are methods for how it is handled and techniques to store, process and manage these diverse data sets.

However there are some engineering challenges; despite developments in technology and platforms to manage these new and unstructured sets, we have some design challenges around how these systems are built and maintained. More so, we can think about the challenges of future-proofing our platforms for the data of tomorrow.

### Image: Cloud Mapping

Cloud computing describes a relatively new development that assists in the supplementing, consumption and delivery of IT services based on the internet. It typically involves the provision of dynamically scalable and often virtualized resources as a service over the internet. It has the potential to alleviate many of these storage and systems we have discussed due to the ability to be provisioned remotely via an internet connection. It is also dynamic in the sense that memory and CPU storage limitations are now mediated via remote access to a server through an internet connection. It also has the ability to be “on demand” through third-party applications which can process and store large datasets via third-party data centres, when requests are made. This saves on maintenance and management in-house of IT systems for most businesses, and as data needs change these requests can be scaled up, or scaled down.

In effect, cloud computing can also allow for:

- Data being processed elsewhere, eliminating the need for systems to accommodate high-level processing (CPU and Memory capacity)
- Space and storage becomes in effect “infinite” as storage is distributed remotely
- Economies of scale benefits, as you only pay for what you use. Subscription services to third-party providers of cloud-based storage and computing platforms are generally the norm. Infrastructure costs are therefore owned by the 3<sup>rd</sup>

- A heavier reliance on internet access, which shifts the issues of capacity from storage to bandwidth.

It is important to remember and be mindful that Cloud Computing is not simply networked computing. Applications and data are not confined to a single computer (they are distributed and can encompass multiple servers and involve multiple companies). It is also not traditional IT outsourcing, where subcontracting occurs. It is also not a buzzword. This is a concept that has now been around and in commercial use (viable) for the past few years.

## The Services of The Cloud

So what can we do with the Cloud?

Many models have emerged for the ways in which this technology can be exploited, and many are relevant for our understanding of Big Data and Analytics practices.

### XAAS/EAAS

“Everything as a Service”

This basically refers to an increasing number of services that are delivered over the Internet rather than be provided locally or on-site. It allows for organisations to call up software components (applications) over a network for use by a business. This can range from email and administration operations, to computing and processing applications. It is technically a subset of cloud computing, with the most common example being Software as a Service (SaaS). However, this “as service” terminology has been associated with many other functions, including communication, infrastructure (IaaS) and platforms (PaaS). Sometimes this is referred to as the SPI model (Software, Platform and Infrastructure). All are core components of Cloud Computing.

**Watch: Presentation by Bill Barney, CEO, Reliance Communications (Enterprise) & Global Cloud Xchange**

A comprehensive presentation on the state of Cloud in 2016 and the impact of the Cloud “revolution”. Some nice critical discussion here which is always appreciated for perspective! Gets a bit techie and industry-focused towards the end, but interesting nonetheless.

---

## Chapter Navigation

Data Warehouse and Cloud

[5.0 Chapter Overview](#)

[5.1 Storage](#)

[5.2 Data Warehousing](#)

[5.3 Cloud Computing](#)

[5.4 Chapter Summary](#)

## 5.3. Cloud Computing

### Cloud Computing Services

| SaaS (Software as a Service)                                                                                                                                                                                                                                                 | PaaS (Platform as a Service)                                                                                                                                                                                                                     | IaaS (Infrastructure as a Service)                                                                                                                                                           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Applications, typically available via a browser. The most common and first service.<br><br>Examples: <ul style="list-style-type: none"><li>• Google Apps (email etc)</li><li>• Windows/Office Live</li><li>• CRM software delivered online, such as Salesforce.com</li></ul> | A platform that enables application developers to host their services. Also provides the ability to build and deploy cloud applications.<br><br>Examples: <ul style="list-style-type: none"><li>• Microsoft Azure</li><li>• Amazon E2C</li></ul> | Service providers offer infrastructure basically to host data and applications.<br><br>Examples: <ul style="list-style-type: none"><li>• Rackspace</li><li>• Amazon E2C and others</li></ul> |

### Image: Cloud Computing Services

This is a nice flow diagram showing who uses what and how

### Watch: Cloud Computing

A lively presentation- if anything a bit outdated – but he makes some salient points towards the end that are interesting to view in retrospect regarding Cloud potential. A nice hint to strategy also.

### Types of Clouds

As well as the various services “on” the Cloud, there are also various types of Cloud itself, which almost resemble types of access. You may have heard of concepts such as “Hybrid Cloud” and wondered what this means, so this will help clarify these concepts.

| Public                                            | Private                                                        | Hybrid                                                 |
|---------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------|
| This is where Cloud services are typically hosted | Where services are managed and hosted by a single organization | Simply put, there is a mix of public and private cloud |

| Public                                                                                                                                                                                                                                                                             | Private                                                                                                                                                                                                                                           | Hybrid                                                                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (corporate, database, storage etc) are delivered to the client via the internet from a third-party service provider.<br><br>Examples: <ul style="list-style-type: none"> <li>• Amazon Cloud Services Offering</li> <li>• Google</li> <li>• Microsoft</li> <li>• Dropbox</li> </ul> | provided within the organisation (in-house). This allows for less restriction on network bandwidth, greater security and other legal concern.<br><br>Examples: <ul style="list-style-type: none"> <li>• HP Data Centres for Enterprise</li> </ul> | services provided from a mix of public and private clouds (in a hybrid cloud).<br><br>Examples: <ul style="list-style-type: none"> <li>• Sales and email marketing, cloud, in-house ERP, Enterprise Resource Planning services.</li> </ul> |

## What are the benefits to public or private clouds?

Public cloud gives the impression of “infinite” storage resources - namely everything is offered and hosted remotely. There are also less up-front commitments and costs to businesses with public cloud, as well as pay-as-you-go options to scale up if more storage and processing is required.

## Cloud Computing and Storage

In terms of organisational perspective, think about where some key points of resistance have emerged from:

| Executive level                                                                                                     | CFO-level                                                                                                 | CIO-level                                                                                 |
|---------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| A buyer-centric approach to technology where applications are available to purchase or rent. Whenever and wherever! | An approach which consumer technology in a pay-as-you-go model where consumers only pay for what is used. | A comprehensive approach to IT where infrastructure is outsourced and managed externally. |

It is important to remember that for all kinds of reasons — technical, political, social, regulatory, and cultural — cloud computing has taken some time to gain commercial momentum. In terms of Big Data storage, the Cloud has not been as enthusiastically embraced by big businesses and enterprises. However, there are many who believe that some key industry players will soon realise that there is a huge ROI opportunity if they do embrace Cloud. These benefits should be now apparent in terms of solving some of our storage issues. In terms of the potential possibilities of the Cloud, we need to go beyond our basic conceptualisation and think purely about how we promote its function for wider business adoption. This includes:

- A shift from a focus on the virtual capability, to the remote delivery of a service. Whether this is public or private is almost moot; the delivery and consumption of this service is what matters most for data and analytics applications.

- Acknowledge some of the key business issues. There are real concerns over privacy and security of data and these need to be addressed so that the business case is solid and bulletproof to promote uptake.
- Technical gaps in the software being remotely provisioned need to be addressed. Everything from the ability to run analytics at scale in a virtual environment to ensuring information processing and analytics authenticity are issues that need solutions.

## Watch: Global Leaders Faculty

Prof. Adrian Cheek, Professor of Pervasive Computing for City University London discusses innovation in the post-industrial age.

The following case study details details some of the many ways the retail sector is facing challenges with Big Data and analytics adoption. Retailers struggle to join up the dots and turn data analysis to advantage. The power of data is becoming clear to organisations of all shapes and sizes, but many in the retail sector are failing to exploit the benefits of data analysis. Research from eCommera found only 23% of UK retailers could quickly use available data to take business decisions. Nearly 50% of retailers believed their business intelligence tools fell short of their needs, with only 16% saying they were confident that their data analytics tools provided the organisational visibility they needed. Computer Weekly attended Demandware's Xchange event in Miami, and it was clear from the keynote speech that data was going to be a key trend for retailers over the coming year. Demandware president and CEO Tom Ebling said retailers should make data an important part of their business strategy. "The future is going to be so personalised, you'll know the customer as well as they know themselves," he said. Tom Davis, global lead for e-commerce at Puma, said retailers were disadvantaged by failing to gain insight into their customers. "If you can't control your data, you can't move fast enough," he said.

## Case: CROCS

### The loyalty opportunity

Footwear brand Crocs has collected data for some time in the US. But it has not yet managed to tie that data to customer profiles on a global basis. So far, Crocs has used transactional data to drive product assorting and recommendations. But the introduction of a new customer relationship management (CRM) program will give the retailer a "big data view" of customers, helping it identify customers who have a propensity to buy more easily and more readily, as well as identifying loyal customers to create a loyalty program.

"We know the customer who shops in multiple channels has a much greater value and provides us an opportunity for growth as we bring new lines to market and expand the brand," said Harvey Bierman, vice-president of global e-commerce at Crocs. Over the

years, many retailers have implemented loyalty programmes, gaining insights into customers' habits.

You can access the full article below:

[Click here. \(Links to an external site.\)](#)

## Watch: Automotive Traders Find Dealing with Gumtree Easy with Salesforce

If you ignore the Salesforce pitch language, it provides an interesting application of Salesforce, an important and popular SaaS application

---

### Chapter Navigation

#### Data Warehouse and Cloud

[5.0 Chapter Overview](#)

[5.1 Storage](#)

[5.2 Data Warehousing](#)

[5.3 Cloud Computing](#)

[5.4 Chapter Summary](#)

### 5.4. Chapter Summary

There are issues and limitations around data storage. We first looked at how storage systems work: some of the structures and systems that are required for Big Data requirements, with a heavy emphasis on Volume. We also looked at some of the key limitations for storage and analytics, the challenges that currently face business. One of the most important is overload of current IT systems and employees. We then explored data warehouses, data marts and the concepts of OLAP and OLTP for storage, retrieval and processing of data. Technologies have evolved, and so we need better solutions.

Cloud came in as our saviour to some of these storage limitations, and we discussed the fundamental concept of what Cloud Computing really means. Due to the remote availability, application and service potential, as well as important operational and efficiency benefits for businesses, Cloud provides a strong case for suitability to Big Data programs. We moved into the various types of Cloud (public, private and hybrid), as well as the services that are offered and how they are provided by third-party

suppliers. This further assisted us in terms of promoting its function for wider business adoption (as well as some of the hurdles).

## Exercise

Try answering these brief review questions:

Choose one of the examples from the previous section and explore what the service offering to businesses is in terms of Cloud potential.

1. Think specifically how this is vital for Big Data applications: would this help alleviate some of the storage limitations we have discussed previously?
2. Why would businesses be ready to adopt and deploy some of these services? List the key benefits to an organisation from a functional or operational perspective.

Read the Case Study below from Computer Weekly (2014)

[Click here.](#) (Links to an external site.)

Think about the following:

1. What key data storage challenges existed for some of these retailers?
2. Why is there a hesitancy for retailers to adopt some of these analytics applications?
3. What was the value here of Cloud services?
4. How did consumers benefit?

Try answering these brief review questions:

1. Describe the main types of storage structures. How has Big Data caused these structures to change? Summarise the key limitations of traditional storage systems.
2. Justify two arguments for why organisations should focus on storage if they want to expand or implement a large-scale data analytics program. What is likely to occur if storage is neglected?
3. Explain the difference between SaaS, PaaS and IaaS.
4. Why would businesses be attracted to hybrid Cloud offerings? What are some of the limitations of a purely public or purely private cloud network?
5. Why do you think businesses have not fully embraced Cloud computing? Explain with reference to both business operations and Big Data programs more specifically. How can this be overcome?

[Answer here.](#)

## References

Computer Weekly (2014), "Are retailers using data analytics to their advantage?", Caroline Baldwin (Ed.), accessed online at:

<http://www.computerweekly.com/news/2240218550/Are-retailers-using-data-analytics-to-their-advantage> [January, 2016]

Gartner (2015) "Gartner's 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor", accessed online at: <http://www.gartner.com/newsroom/id/3114217> [January, 2016]

Lumidata (2015) "Don't be Overwhelmed By Big Data", accessed online at: [http://www.lumidata.com/sites/lumidata.com/files/whitepapers/Don't%20Be%20Overwhelmed%20by%20Big%20Data\\_2015pdf.pdf](http://www.lumidata.com/sites/lumidata.com/files/whitepapers/Don't%20Be%20Overwhelmed%20by%20Big%20Data_2015pdf.pdf) [January, 2016]

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons. [http://docs.oracle.com/html/E10312\\_01/dm\\_concepts.htm](http://docs.oracle.com/html/E10312_01/dm_concepts.htm) Surajit Chaudhuri & Umeshwar Dayal (1997). "An overview of data warehousing and OLAP technology". SIGMOD Rec. ACM. 26 (1): 65.

## Extra Resources

| Resource                                                                                                             | Description                                                                                                                                                                                                                                                                             |
|----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Light, easy and fun                                                                                                  |                                                                                                                                                                                                                                                                                         |
| Ian Moyse (2012) Vimeo on Cloud Computing:                                                                           | A lively presentation- if anything a bit out of sync but makes some salient points towards the future of computing interesting to view in retrospect regarding the nice hint to strategy also.<br><a href="#">Watch here.</a> (Links to an external site.)                              |
| Salesforce's case study for Gumtree (2014)                                                                           | If you ignore the Salesforce pitch language it's an interesting application of Salesforce, an extremely popular SaaS application.<br><a href="#">Watch here.</a> (Links to an external site.)                                                                                           |
| For the academically curious                                                                                         |                                                                                                                                                                                                                                                                                         |
| Presentation by Bill Barney, CEO, Reliance Communications (Enterprise) & Global Cloud Xchange, Hong Kong SAR China,. | A comprehensive presentation on the state of the industry and the impact of the Cloud "revolution" and the discussion here which is always appreciated. Gets a bit techie and industry-focused but interesting nonetheless.<br><a href="#">Watch here.</a> (Links to an external site.) |
| Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on   | The most comprehensive plain-language overview of Cloud and Big Data and its complex relationship                                                                                                                                                                                       |

| Resource                                                                           | Description                                                                              |
|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| cloud computing: Review and open research issues. Information Systems, 47, 98-115. | interesting conclusions, and excellent overview of the key terminology we have discussed |
|                                                                                    | <a href="#">Download here.</a> (Links to an external site)                               |

---

## Chapter Navigation

### Data Warehouse and Cloud

[5.0 Chapter Overview](#)

[5.1 Storage](#)

[5.2 Data Warehousing](#)

[5.3 Cloud Computing](#)

[5.4 Chapter Summary](#)

---

## Chapter 6

### 6.0. Chapter Overview - Processing Platforms for Data Discovery

This chapter explores how businesses manage and harness large databases and process them to discover dynamic and real-time information. The chapter introduces Hadoop and HDFS, the common standard for data storage, The Hadoop EcoSystem, and MapReduce. Furthermore, it looks into emerging processing technologies (such as Cassandra) and contemporary solutions to data processing issues.

1. Hadoop.
  2. MapReduce.
  3. Cassandra.
- 

## Chapter Navigation

## Processing Platforms for Data Discovery

### [6.0 Chapter Overview](#)

### [6.1 Hadoop](#)

### [6.2 MapReduce](#)

### [6.3 Cassandra](#)

### [6.4 Chapter Summary](#)

## 6.1. Hadoop

### From Our Last Chapter

Storage is a critical element for Big Data. Data, as we know, has to be stored somewhere, readily accessible and protected. This has proved to be an expensive challenge for many organisations, since network-based storage (not forgetting Cloud systems) can be very expensive to purchase and manage. Storage has evolved to become one of the more “standard” elements in a typical data center. Indeed, storage technologies have started to approach a critical commodity-like status for businesses. Last chapterspoke about some of the evolving data needs that can put the strain on storage technologies. Traditional storage technologies are not suited to deal with the terabytes (and petabytes!) of unstructured information presented by Big Data as we have discussed previously. Success with Big Data analytics demands something more: a new way to deal with large volumes of data and a new storage platform approach.

### Image: Rise of Distributed Data Systems

The ability to design, develop, and implement a big data program is directly dependent on an understanding of the underlying computing platform, both from a hardware and, more importantly, from a software perspective.

Because simplistic traditional computer systems are limited in capacity, they cannot easily accommodate massive amounts of data. High-performance platforms are composed of collections of computers in which the massive amounts of data and requirements for processing can be distributed among a pool of resources, easing this burden. The terms distributed database and distributed processing are closely related, yet have distinct meanings.

| Distributed database                                                                                | Distributed processing                                                                                  |
|-----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| A set of databases in a distributed system that can appear to applications as a single data source. | The operations that occur when an application distributes tasks among different computers in a network. |

For example, a database application typically distributes front-end presentation tasks to client computers and allows a back-end database server to manage shared access to a database.

### Watch: Centralised vs. Distributed Databases

Please watch this short video which canvasses the key ways centralised databases differ from distributed databases.

### Read: The Year of Data Walls and Cloud Battles

A highly recent and quite accessible article on the year that was 2015 and some nice predictions on the Big Data and Hadoop space for 2016 and beyond.

[Read here.](#) (Links to an external site.)

### The Hadoop File System (HDFS)

The Big Data foundation is composed of two major systems. The first stores the data and the second processes it. Big Data storage is often synonymously interchanged with the Hadoop File System (HDFS), but traditional data warehouses can also house Big Data. HDFS is distributed data storage that has become the de facto standard because you can store any type of data without limitations on the type or amount of data. One of the reasons HDFS has become so popular is that you don't have to do any set up to store the data. In traditional databases, you need to do quite a bit of "set up" to store data.

Data that will be housed in the database is set up by creating a schema. The schema is the blueprint for how data is placed into tables with columns. The schema also contains rules about the data being stored — for example, which column(s) will be used to find or index the data in a particular table. With HDFS, you don't set up a blueprint: you simply dump the data into a file. (Think about it as cutting and pasting all your data into a huge Microsoft Word file!) This approach makes sense for massive quantities of data when the value of that data is unknown. The name "Hadoop" was given by one of Doug Cutting's sons to that son's toy elephant. Doug used the name for his open source project because it was easy to pronounce and to Google.

An interview with Doug Cutting discussing more can be found below:

[Click here](#) (Links to an external site.)

### So Why Is It So Good?

There are many Big Data technologies that have been making an impact on the new technology stacks for handling Big Data, but the HDFS seems to be the most prominent and more and more businesses are just now starting to leverage its capabilities. When

data lands in the cluster, HDFS breaks it into pieces and distributes those pieces among the different servers participating in the cluster. Each server stores just a small fragment of the complete data set, and each piece of data is replicated on more than one server. HDFS is also used to complement traditional data warehouses because of limitations on the type of data the database supports, and the size of data that the database can store. Often, traditional databases support the data type but make it impractical to manipulate the data once it is stored thereby rendering it fairly useless.

Big Data processing involves the manipulation and calculations that occur on the Big Data. Traditional databases have differing abilities to process Big Data sets effectively and efficiently. The Hadoop platform is designed to solve problems caused by massive amounts of data, especially data that contain a mixture of complex structured and unstructured data, which does not lend itself well to being placed in tables. Hadoop works well in situations that require the support of analytics that are deep and computationally extensive. In some ways, it can be said to perform similar to a supercomputer.

### Quote

“For the decision maker seeking to leverage Big Data, Hadoop solves the most common problem associated with Big Data: storing and accessing large amounts of data in an efficient fashion. The intrinsic design of Hadoop allows it to run as a platform that is able to work on a large number of machines that don’t share any memory or disks. With that in mind, it becomes easy to see how Hadoop offers additional value: Network managers can simply buy a whole bunch of commodity servers, slap them in a rack, and run the Hadoop software on each one.”

Frank Ohlhorst, 2015

### Watch: Data Science at the New York Times

An excellent wide-ranging keynote by Chris Wiggins on how the media space has been transformed by data, specifically Hadoop applications. I love the tone, and some of the very interesting ways in which data is mined and monetised for readership. Fascinating!

### Image: How Did We Get Here?

Hadoop originated from the “Nutch” open-source project on search engines in the early 2000s. Google revolutionised the field of Big Data and Cloud Computing in 2004 when they released a cache of papers describing how they handled these topics. In 2006, Lucene started a subproject using operations from this project, but it wasn’t until 2008 when Apache took over Hadoop and made it a top-level project when it started to “take off” in terms of computational ability and feasibility.

Who is now using Hadoop to manage their Big Data? Hadoop is now arguably the backbone of:

- Amazon/A9
- Facebook
- Google
- IBM
- Disney
- New York Times
- Yahoo!
- Twitter
- LinkedIn

## How Can We Use Hadoop

Hadoop has a variety of uses beyond a distributed storage platform. The following detail the many different ways it has revolutionised information management across a variety of sectors.

- Large dataset analysis (eg: AOL data warehousing)
- Social Media analysis
- Text mining and patterns search (e.g.: recommendations systems - Facebook and Disney)
- Machine Log analysis
- Geo-spatial analysis
- Trend Analysis (e.g.: search analysis - Yahoo, Amazon)
- Genome Analysis
- Drug Discovery
- Fraud and Compliance Management
- Video and Image Analysis (e.g.: Facebook, NY Times)
- Output analysis (e.g.: computing carbon footprint, emissions etc)

### Watch: Hadoop in Business Advancement

Watch this video to see how Hadoop is advancing video and image analysis and changing the 'face' of emotional recognition and see what it means for businesses.

---

## Chapter Navigation

Processing Platforms for Data Discovery

[6.0 Chapter Overview](#)

[6.1 Hadoop](#)

[6.2 MapReduce](#)

[6.3 Cassandra](#)

[6.4 Chapter Summary](#)

Immersive Reader

## 6.2. MapReduce

### Processing Large Datasets

Because Hadoop stores the entire dataset in small pieces across a collection of servers, analytical jobs can be distributed, in parallel, to each of the servers storing part of the data. Each server evaluates the question against its local fragment simultaneously and reports its results back for collation into a comprehensive answer. MapReduce is the agent that distributes the work and collects the results. It does this by automatically dividing the processing workload into smaller workloads that are distributed.

MapReduce is a programming model for processing large data sets on distributed computing. MapReduce has emerged as the de-facto processing software for HDFS, and is a fault-tolerant parallel programming framework designed to harness Big Data processing capabilities.

When working with HDFS, data manipulation and calculations are programmed using the MapReduce framework in the language of choice, which is typically Java programming language. Java is a general-purpose computer programming language that is specifically designed to have as few implementation dependencies as possible. It is intended to let application developers “write once, run anywhere”, meaning that compiled Java code can run on all platforms that support Java without the need for recompilation. Java is one of the most popular programming languages in use.

### Parallel Computing Framework

A parallel computing framework may be a bit meaningless to most of us. The following analogy helps to explain this concept and how MapReduce operates. For example, a factory with 10 assembly lines receives an order to create 500 toy trucks. One assembly line could create all 500 toy trucks or, alternatively, there could be a division of labor among the assembly lines where each assembly line produces 50 toy trucks. If each assembly line started at the same time and everything went flawlessly each assembly line would complete their production of 50 trucks simultaneously. This efficient division of labor was fairly straightforward because each toy truck could be produced independently.

However, if three of the assembly lines could only produce the engine and the remaining seven assembly lines could only produce the balance of the toy truck the division of labor becomes more complicated. In this scenario, planning has to take into account that there is a dependency between the engine production and rest of the toy truck production. Just like in the toy truck production, some data manipulation and calculations can be performed independently. To maximize the processing throughput, MapReduce assumes that the workloads being distributed are independent tasks and the workload is equally divided just like the division of labor to the 10 assembly lines to produce 50 trucks each. However, if there are dependencies in the processing workload, the MapReduce framework is unaware of those dependencies. The programmer has to be aware of those dependencies and has to specifically divide the workload up in the program understanding that MapReduce will automatically distribute tasks. This type of programming is called parallel programming.

Just like the production planning that has to occur in dividing the workload between assembly lines that only produce engines and all the other assembly lines was more complicated, parallel programming is more complicated. One of the benefits of MapReduce and some data warehouse appliances is that the easier independent processing is automatically handled by the framework or data warehouse appliance.

(Minelli, Chambers, & Dhiraj 2012)

### Image: Map. Shuffle. Reduce

It is helpful to think about MapReduce as an “engine” because that is exactly how it works. You provide input (fuel), the engine converts the input into output quickly and efficiently, and you get the answers you require. Hadoop MapReduce includes several stages, each with an important set of operations helping to get to your goal of getting the answers you need from big data. The process starts with a user request to run a MapReduce program and continues until the results are written back to the HDFS.

The basic stages include:

| Mapping                                                                                        | Shuffling                                                                                                                 | Reduce                          |
|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| Writing the output files to temporary storage, assigning all of the input files to processing. | Data is redistributed according to the mapped function. All similar data is sorted and assigned to the reduce processors. | Processing now occurs parallel. |

The MapReduce system then collects all the Reduce output data, and it is sorted to produce a final outcome.

### Image: Another Way of Looking at Things

This highlights our Big Data journey, right across and through the Hadoop processing system (HDFS). We can see data being distributed, assigned and processed from large daunting sets of data, to manageable, actionable groupings. It is a process of sorting in order to compute and it results in lightening-fast computing ability that is practically and virtually highly efficient.

### Image: The Data Processing Landscape

You might see or hear the term “Apache Hadoop” in the literature and online readings. This is because the software is managed by an open-sourced foundation called the Apache Software Foundation. They are an American membership-based organisation that are decentralised in structure, but they also oversee the many components of the Hadoop framework.

The Hadoop framework is composed of the following chapters:

### Image: The Hadoop Ecosystem

The term Hadoop has come to refer not just to the base chapters above, but also to the ecosystem, or a collection of additional software packages that can be installed on top of or alongside Hadoop.

| Chapter                               | Description                                                                                                                           |
|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| Hadoop Common                         | Contains libraries and utilities needed by other Hadoop chapters.                                                                     |
| Hadoop Distributed File System (HDFS) | A distributed file-system that stores data on commodity machines, which we have covered previously.                                   |
| Hadoop YARN                           | A resource-management platform responsible for managing computing resources and using them for scheduling of users' applications.     |
| Hadoop MapReduce                      | An implementation of the MapReduce programming model for large scale data processing that we have previously covered.                 |
| Package                               | Description                                                                                                                           |
| Hive                                  | A distributed data warehouse for data that is stored in HDFS.                                                                         |
| Hbase                                 | Is a columnar database and an implementation of Google BigTable- highly scalable to millions of rows.                                 |
| Flume                                 | Apache Flume is a distributed, reliable and highly available service for efficiently collecting and moving large amounts of log data. |
| Mahout                                | A library of machine learning statistical algorithms implemented in MapReduce and Hadoop.                                             |

| Package   | Description                                                                                                                                                                                                                |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sqoop     | Used for transferring bulk data between Hadoop and traditional structured data platforms.                                                                                                                                  |
| Pig       | A platform for the analysis of very large data sets that runs on HDFS and with an interpreter consisting of a compiler that produces sequences of MapReduce programs and a language of the query language named Pig Latin. |
| Zookeeper | A simple interface to the centralised coordination of services (such as naming, configuration, and synchronisation) used by distributed applications.                                                                      |

## Image: The Data Processing Landscape

And it doesn't stop there. Hadoop is only one of the many providers of processing services (albeit they are open-source). There are cross-infrastructure platforms provided by IBM and SAP, as well as applications provided by Google, Microsoft and other private analytics and infrastructure businesses. The image on the left perhaps outlines the complexity of the analytics processing marketplace! The following video from Cloudera is another example of another such provider.

## Watch: Different Platforms for Different Sectors

The Wall Street Journal sat down with Cloudera Chief Executive Tom Reilly to talk about the different ways in which Hadoop is being deployed by certain businesses. These applications are revealing in the sense that Cloudera has its own market positioning and unique value proposition to certain sector groups.

## Choosing The Right Platform

Current forecasting suggests that the Big Data technology and services market will grow at beyond 40% per year, which is over seven times the estimated growth rate for the overall IT and communications business (Framingham, 2019). By the end of this year alone, it is predicted that 75% of all data on the planet will be residing in Hadoop systems. While Hadoop is applicable to the majority of businesses, it is not the only provider, especially when it comes to open-source platforms. Once an organisation has decided to leverage its massive data sets, setting up the infrastructure will not be the biggest challenge. The biggest challenge may come from deciding to go it alone with an open source provider, or to turn to one of the commercial implementations of Big Data technology. Providers such as Cloudera, Hortonworks, and MapR are commercialising Big Data technologies, making them easier to deploy and manage.

Add to that the growing crop of Big Data on-demand services from cloud services providers, and the decision process becomes that much more complex. Decision makers will have to invest in research and perform due diligence to select the proper platform and implementation methodology to make a business plan successful. However, most of that legwork can be done during the business plan development phase, when the pros and cons of the various Big Data methodologies can be weighed

and then measured against the overall goals of the business plan. Which technology will get us there the fastest, with the lowest cost, and without sacrificing future capabilities?

## Moving on from Hadoop

Hadoop and the world of big data have transitioned from being buzzwords and hype to being a reality. Businesses have realised the value of Big Data and are beginning to understand the use cases for the technologies in the Hadoop ecosystem. While the penny may have dropped for some business, Hadoop has not yet forged a name for itself as being the most simplistic set of technologies to use. Some confusion remains about what exactly it is and this adds to the complexity. Hewlett-Packard proposed a new kind of computing architecture that focuses on memory, as opposed to processing power, as the center of a computing system in 2015. Around the same time, IBM, which has championed Hadoop and put it at the center of its big data strategy, announced it was working on a faster data-processing engine, called Spark. Additionally, a senior executive at Cloudera, probably the largest Hadoop company, hinted that they are prepared to see key parts of Hadoop diminish in importance, and was increasingly distributing Spark.

Hadoop is complex technology but an important thing to remember is that the business doesn't necessarily need to know how complex it is or how it works. The work done by the Apache Hadoop developer community along with the large vendors has helped to mask the complexities so that businesses can instead focus on putting it to use. Although the technology is complex it is not something that is fundamental to the user being able to use it. The important elements are knowing your use cases and having best practices in place in order to make sure it benefits the business. (Nunns, 2016)

### Read: Companies Move On From Big Data Technology Hadoop

If you are interested in future tech and directions of this space, this furthers the discussion around the current state of the Hadoop market and is a nice overview of what some of the big key players are doing in the technology space from 2015 onwards

[Read here. \(Links to an external site.\)](#)

---

## Chapter Navigation

Processing Platforms for Data Discovery

[6.0 Chapter Overview](#)

[6.1 Hadoop](#)

[6.2 MapReduce](#)

[6.3 Cassandra](#)

[6.4 Chapter Summary](#)

## 6.3. Cassandra

### Beyond Hadoop

Another name in the Big Data realm is the Cassandra database, a technology that can store 2 million columns in a single row. Cassandra is ideal for appending more data on to existing user accounts without knowing ahead of time how the data should be formatted. It is an open-source system managed by Apache. Cassandra's origins can be traced to Facebook, which needed a massive distributed database to power the service's inbox search. Facebook wanted to use the Google Bigtable architecture, however, Bigtable had a serious limitation in that it depended on a single node to coordinate all read-and-write activities on all of the nodes. This meant that if the head node went down, the whole system would be useless.

Cassandra was built on a distributed architecture called Dynamo. Amazon uses Dynamo to keep track of what customers are putting in their shopping carts. Dynamo gave Cassandra an advantage over Bigtable, as Dynamo is not dependent on anyone master node. Any node can accept data for the whole system, as well as answer queries. Data is replicated on multiple hosts, creating resiliency and eliminating the single point of failure. Google Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Bigtable, (from URLs to web pages to satellite imagery).

(Google Inc, 2016; Ohlhorst, 2012)

---

## Chapter Navigation

### Processing Platforms for Data Discovery

[6.0 Chapter Overview](#)

[6.1 Hadoop](#)

[6.2 MapReduce](#)

[6.3 Cassandra](#)

[6.4 Chapter Summary](#)

## 6.4. Chapter Summary

By now you should have a deeper understanding not only of the way enterprises and large organisations manage and harness these databases, but also how they enable the processing and crunching of dynamic and real-time information. We looked at the importance of processing, and how distributed data systems work. This is the fundamental concept behind the Hadoop system (HDFS) and we unpacked its various uses in terms of managing large datasets and the development of storage processes. From that we delved into MapReduce - a critical processing platform that makes this seemingly infinite data pool so much more efficient through its Map/Shuffle/Reduce framework.

This chapter also touched upon the wider Hadoop ecosystem, the various other tools and platforms that exist in this processing universe. While we did not go into a computation exercise specifically, what is important here is the application of these tools and the importance in selecting the right analytics systems. Again, we looked at some of the drivers behind Hadoop incorporation in business, as well as some emerging alternatives in the private marketplace beyond open-source offerings. These contemporary solutions to Big Data processing problems allow us to imagine a future of further analytic and informatics possibility – how will these developments assist us with Big Data management, storage and processing? Furthermore, we need to consistently evaluate and ask ourselves which technology will get us there the fastest, with the lowest cost, and without sacrificing future capabilities.

### Question Time

Try answering these review questions:

1. What are the key benefits of a distributed data system? How does it differ from a modular, centralised system?
2. Why did Hadoop become so widely accepted? What are its benefits over traditional storage and processing platforms?
3. Explain how an organisation like Facebook might use Hadoop in its Big Data processing and management.
4. Why is the Map/Shuffle/Reduce process of MapReduce so efficient? Why would this benefit businesses in terms of data outcomes?
5. Describe two different tools in the Hadoop ecosystem and why are they have might have benefits to two different industries.

[Answer here.](#)

## Glossary

### Computation and Data

Big Data as an industry is quite dynamic, and still is a degree of flux....as you have probably ascertained, there are new terms and technology appearing almost weekly! This fast-paced environment is fuelled by open-source communities, emerging tech companies, and key industry players such as IBM, SAP and Oracle. The following should help with some of the terminology!

| Terminology                     | Description                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Analytics                       | Using maths to derive meaning from data                                                                                                                                                                                                                                                                                                        |
| Analytics Platform              | A set of analytic tools and computational power used to query and analyze data                                                                                                                                                                                                                                                                 |
| Appliance                       | Optimised hardware and software purpose built for a specific set of tasks                                                                                                                                                                                                                                                                      |
| Batch                           | A job or process that runs in the background without human intervention                                                                                                                                                                                                                                                                        |
| Big Data                        | The de facto standard definition of big data is data that goes beyond the limits of data along three dimensions: volume, variety, velocity. (It's also four dimensions!) The combination of these dimensions makes the data complex to ingest, process, and visualise.                                                                         |
| Cloud                           | General term used to refer to any computing resources — software or hardware — that is delivered as a service over a network.                                                                                                                                                                                                                  |
| Columnar Database               | The storing and optimising of data by columns. Particularly useful for processing that uses data based on a column.                                                                                                                                                                                                                            |
| Data Mining                     | The process of discovering patterns, trends, and relationships from large datasets using machine learning.                                                                                                                                                                                                                                     |
| Distributed Processing          | The execution of a program across multiple CPUs.                                                                                                                                                                                                                                                                                               |
| Grid                            | Loosely coupled servers networked together to process workload                                                                                                                                                                                                                                                                                 |
| Hadoop                          | An open-source project framework that can store large unstructured data and processes the unstructured data (MapReduce) in a cluster of servers                                                                                                                                                                                                |
| HDFS                            | The Hadoop File System, which is the storage mechanism for Hadoop                                                                                                                                                                                                                                                                              |
| MapReduce                       | A parallel computational batch processing framework for Hadoop where the job is written in Java. The job breaks up a larger problem into smaller problems and distributes the workload across the grid so that jobs can be worked on simultaneously (mapper). A master job (reducer) collects all the intermediate results and combines them.  |
| Real-Time                       | Today it is colloquially defined as immediate processing. Real-time processing originated in 1950s when multi-tasking machines provided the capability to interrupt a task for a higher priority task to be executed. These types of real-time systems were used in space program, military applications, and many commercial control systems. |
| Relational Database             | The storing and optimising of data by rows and columns.                                                                                                                                                                                                                                                                                        |
| Semistructured Data             | Unstructured data that can be put into a structure by available for processing                                                                                                                                                                                                                                                                 |
| SQL (Structured Query Language) | The language for storing, accessing, and manipulating data in a relational database                                                                                                                                                                                                                                                            |

| Terminology       | Description                        |
|-------------------|------------------------------------|
| Structured Data   | Data that has a pre-set format.    |
| Unstructured Data | Data that has no preset structure. |

## References

CNBC (2013) “Big Data Giveaway’s Big Payoff”, accessed online at: <http://video.cnbc.com/gallery/?video=3000165622> [February, 2016]

Framingham, M. (2019) ‘IDC Forecasts Revenues for Big Data and Business Analytics Solutions will Reach \$189.1Billion this Year with Double-digit Annual Growth Through 2022’, International Data Corporation, [Online]  
<https://www.idc.com/getdoc.jsp?containerId=prUS44998419>

Google (2016) “BigTable” accessed online at: <http://static.googleusercontent.com/media/research.google.com/en//archive/bigtable-osdi06.pdf> [February, 2016]

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Nunns (2016) “How and when to use Apache Hadoop to bring big data value to business”, CBR Online, accessed online at: <http://www.cbronline.com/news/big-data/software/how-and-when-to-use-apache-hadoop-to-bring-big-data-value-to-business-4807141> [February, 2016]

Oehlhorst, F. J. (2012). Big data analytics: turning big data into big money. John Wiley & Sons.

Oracle (2016) “Database Administrator’s Guide”, accessed online at: [https://docs.oracle.com/cd/B10501\\_01/server.920/a96521/ds\\_concepts.htm](https://docs.oracle.com/cd/B10501_01/server.920/a96521/ds_concepts.htm) [February, 2016]

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge.

## Extra Resources

| Resource                                           | Description                               |
|----------------------------------------------------|-------------------------------------------|
|                                                    | Light, easy and fun                       |
| Muglia, B (2016) “The Year of Data Walls and Cloud | A highly recent and quite accessible arti |

| Resource                                                                                                                                                                                           | Description                                                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Battles”, Huffington Post.                                                                                                                                                                         | was 2015 and some nice predictions on Hadoop space for 2016 and beyond.<br><a href="#">Read here. (Links to an external site.)</a>                                                                                                                   |
| Hardy, Q (2015) “Companies Move On From Big Data Technology Hadoop”, NY Times                                                                                                                      | Similar to the above, but far more optimistic about machine learning and its links to big data. It can be aided in the future, as well as some of the futurism!<br><a href="#">Read here. (Links to an external site.)</a>                           |
| Wiggins, C (2015) “Data Science At The New York Times”                                                                                                                                             | An excellent wide-ranging keynote by C. Wiggins. The media space has been transformed by Hadoop I love the tone, and some of the ways in which data is mined and monetized. Fascinating!<br><a href="#">Watch here. (Links to an external site.)</a> |
| For the academically curious                                                                                                                                                                       |                                                                                                                                                                                                                                                      |
| Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. Journal of Big Data, 2(1), 1-36. | <a href="#">Read here. (Links to an external site.)</a>                                                                                                                                                                                              |

---

## Chapter Navigation

### Processing Platforms for Data Discovery

#### [6.0 Chapter Overview](#)

#### [6.1 Hadoop](#)

#### [6.2 MapReduce](#)

#### [6.3 Cassandra](#)

#### [6.4 Chapter Summary](#)

---

## 7.0. Chapter Overview - Business Intelligence

This chapter introduces the concept of Business Intelligence and how it relates to Big Data analytics. It introduces data mining for enabling insights by setting real-time analytics direction. Current methods of analysing data including (not limited to) trend analysis, sampling, text mining, sentiment analysis, web analytics, and graphs are examined, with examples of visualisation.

1. Analytics.
2. Data Mining.
3. Modelling and Methods.

### Quote

“One thing is clear: Every enterprise needs to fully understand big data - what it is to them, what it means to them - and the potential of data-driven marketing, starting today. Don't wait. Waiting will only delay the inevitable and make it even more difficult to unravel the confusion.”

Forbes, 2013

---

## Chapter Navigation

**Business Intelligence**

[7.0 Chapter Overview](#)

[7.1 Analytics and Business Intelligence \(BI\)](#)

[7.2 Data Mining](#)

[7.3 Modelling and Methods](#)

[7.4 Chapter Summary](#)

## 7.1. Analytics and Business Intelligence (BI)

### Analytics

#### Understanding Big Data

We have used the terms Big Data and Analytics almost interchangeably throughout the course. Now that we have grasped a working understanding of Big Data, it is time to pick apart the ranging notion of analytics as we start to look at what we do with this data. Analytics is another broad term that is used to define an entire spectrum of data computations and the patterns that can be found in these datasets. Information resulting from the systematic analysis of data or statistics.

The colloquial definition of analytics is using past information to provide insights into the future. Now, if we combine massive amounts of data (a.k.a. Big Data) along with powerful analytics (a.k.a. advanced analytics) there is a greater deal of texture or context to predict future outcomes, which allows us to then take it to the next level and use those insights to automate actions or to provide a short list of actions that are likely to result in significantly improved outcomes.

Business intelligence (BI) represents the tools and systems that play a key role in the strategic planning process within a corporation. These BI systems allow a company to gather, store, access and analyse corporate data to aid in decision-making. Generally these systems will illustrate business intelligence in the areas of customer profiling, customer support, market research, market segmentation, product profitability, statistical analysis, and inventory and distribution analysis to name a few. (Webopedia, 2016)

### Watch: Global Leaders Faculty

Mr Dominic Barton, Managing Director of McKinsey & Co, explores how technology can be an enabler.

## BI and Big Data

Many of these techniques started in statistics and were used in Business Intelligence dashboards and scorecards. Business Intelligence helps us aggregate information about events that took place in the past. As valuable as BI is, it essentially provides us with a look in the rearview mirror. Many organisations and individuals are confused by the differences introduced by Big Data, especially the differences between Business Intelligence (BI) and Big Data science. Big Data Science is not big BI. Big Data is a key enabler of a new discipline called data science that seeks to leverage new sources of structured and unstructured data, coupled with predictive and prescriptive analytics, to uncover new variables and metrics that are better predictors of performance.

That said, BI and data science share many of the same objectives:

- Extracting value from data
- Cleansing dirty data
- Transforming and aligning data
- Helping support decision making

The questions, characteristics, processes, tools, and models are actually quite different.

(Schmarzo, 2015)

### Image: What BI is Looking For

BI focuses on descriptive analytics: that is, the “What happened?” types of questions.

Examples include:

- How many X did I sell last month?
- What were sales by zip/post code for Christmas last year?
- How many units of X were returned last month?
- What were company revenues and profits for the past quarter?
- How many employees did I hire last year?

BI focuses on reporting on the current state of the business, or as is now commonly called Business Performance Management (BPM). BI provides retrospective reports to help business users to monitor the current state of the business and answer questions about historical business performance. These reports and questions are critical to the business, sometimes required for regulatory and compliance reasons. BI can apply some rudimentary analytics (time series analysis, previous period comparisons, indices, shares, and benchmarks) to help business users to flag areas of under- and over-performance. But even these analytics are focused on monitoring what happened to the business.

### From Reporting to Actions

Various organisations across service and product-based industries have flawlessly executed on developing analytics but have failed miserably on integrating these with actions. Creating analytics does not automatically result in institutionalising analytics. To put it another way, the difference between doing analytics versus being analytical. While doing analytics focuses on the creation of analytics, being analytical balances and integrates creation of analytics with consumption of analytics. Organisations able to bridge this development-integration gap will be able to capitalise on analytics as a source of competitive advantage.

Indeed, an ongoing key challenge that companies face is: How to convert thought into action and bridge the gap between analytics creation and consumption?

### Watch: IBM Big Data and Analytics

This IBM YouTube video looks at the value of analytics with respect to the banking sector. Think about the ways that customer profitability is able to be assessed from the

bank’s perspective, but also think about some of the benefits to the customer in terms of service given. Take a note also of the many sources of data here – you will be surprised at the scope of analytics available across a variety of devices!

What is Listening?

As data gets larger, it becomes increasingly difficult to fully grasp the meaning and magnitude of the data through exploratory analysis without effective visualisation tools and experience at discerning between the noise (Big Data) and the signal in the noise. Once you are comfortable with the noise, you can become adept at identifying the signal more readily. Finding signals can be like finding a needle in a haystack. The sheer volume of data could distort or mask the signal or it could be that there is a long tail and the bulk of the signal is in a large curve or the emerging signal is in the smallest area.

Signals can also be cyclical and only emerge when viewed over a long period of time. Many economic cycles take 10 years to unfold. So, if you’re only looking at three years of data you won’t see the signal or may misinterpret the signal. Some signals are complex and require an understanding of several related factors in order to be identified. This is where Big Data visualisation tools come into play, making it easier to wrap your arms around unwieldy data. How does the biggest wireless carrier in the United States keep their customers happy? All 100 million of them? Provide the best network with the best customer experience and listen.

Watch: How did Verizon Wireless listen using Big Data and analytics?

- How does the biggest wireless carrier in the United States keep their customers happy?
- All 100 million of them?
- Provide the best network with the best customer experience and listen. How did Verizon Wireless listen using Big Data and analytics?

The Listening Process

The following provides a framework for streamlining the analytics process and makes it easier to transfer throughout an organisation. Each of the stages in the cycle is key to enterprise synthesis of the analytics, and enables movement towards the process of listening, rather than simply collecting and creating more data.

|               |                                                                                                                                                                                                                                                               |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Communication | Successful consumption of analytics is a collaborative endeavour. The first step is to take your analytics intent beyond your core team and sell it to a wider group of stakeholders—your prospective daily consumers of analytics in your organisation.      |
| Implement     | Organisations that successfully consume analytics are driven by leadership vision and consensus in the organisation and allows for moving ahead without the need for buy-in on board, every step of the way. The initial focus of implementation should be on |

|         |                                                                               |
|---------|-------------------------------------------------------------------------------|
|         | ingredients in place to create the basic human and technology infrastructures |
| Measure | The true test of consumption is to use analytics to measure itself.           |

Be forewarned against overestimating the impact of analytics by discounting the human element. A successful business decision is a healthy combination of business experience and analytics—both merit equal credit for successful analytics consumption.

|                           |                                                                                                                                                                                                                                                                                                                   |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Align Incentives          | Successful consumption of analytics mandates creation of more structured d processes, which are driven by data and analysis. This puts constraints on fr experience-driven decision making. Incentives need to be carefully managed outcomes. More on this in upcoming chapters.                                  |
| Develop Cognitive Repairs | Our everyday decision-making is influenced by numerous biases intrinsic to h Presence of data and analysis challenge biases and drive us out of our comfor counter-intuitive business insights based on data emerge, leaders in the organa it right for all to see, in order to expose biases and create repairs. |

It is necessary to acknowledge the holistic nature of listening to data and its organisation-wide influence.

### Image: Listening to Solve Problems: Modelling

Data modeling is often the first step in database design and object-oriented programming. Designers first create a conceptual model of how data items relate to each other. Data modeling involves a progression from conceptual model to logical model to physical schema.

Verhoef, P. C., Kooge, E., & Walk, N. (2016)

However, Big Data presents a problem that needs to be solved in the data modeling approach (above).

The difference with the problem-solving approach is that the focus is more on data, especially the use of new data sources. Analysts focus too much on the data and lack strong conceptualisations on why specific relationships in the data are found. Outcomes can sometimes show spurious correlations in which the association is based on an underlying, unobserved variable. The approach can be overly data-driven instead of problem-driven and can easily become a data-mining exercise. However, an advantage over the problem-solving approach is that the data-modeling approach is more flexible in terms of data usage. This has the potential to result in more innovative model solutions.

**Read: What Cancer Researchers can learn from direct-to-consumer companies (2017) Harvard Business Review.**

Organisations striving to find new ways to attack cancer have much to learn from direct-to-consumer (DTC) companies. Specifically, they can profit from DTC firms' expertise in persuading their customers to provide and share their data. This is something many cancer patients don't do because they are unaware of the data's importance or their power to instruct institutions to share it

[Read here. \(Links to an external site.\)](#)

**Watch: Global Leaders Faculty**

Prof. Adrian Cheek, Professor of Pervasive Computing for City University London, covers technology, government and society.

---

## Chapter Navigation

[Business Intelligence](#)

[7.0 Chapter Overview](#)

[7.1 Analytics and Business Intelligence \(BI\)](#)

[7.2 Data Mining](#)

[7.3 Modelling and Methods](#)

[7.4 Chapter Summary](#)

## 7.2. Data Mining

### So What is Data Mining?

Data mining is a much more explorative analytical strategy. Typically, data is not predefined and a defined problem does not guide the analysis. Also, no hypotheses are implicitly or explicitly stated. Scientifically, it is a typical inductive analysis – that is, hoping to generate a new theory from the data. This means the key belief is that the widely available data can provide valuable insights just by digging into the data. In doing so, relevant new, and potentially valuable, relationships can be discovered. Data Mining is an analytic process designed to explore data (usually large amounts of data – essentially Big Data) in search of consistent patterns and/or systematic relationships between variables. Findings are then validated by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction. Predictive data mining is the most common type of data mining and one that has the most direct business applications. More on this coming later in the next chapter.

(Dell, 2015)

## How to Mine

Data mining is an iterative and interactive process. There are multiple stages of processing and computation which are repeated. In the data mining process results are interpreted via validation methods, before these large data sets are converted into working knowledge. While repetitive in process, this approach can be described as interactive as it looks to involve many actors as possible, from data scientists to CIOs to informatics and BI teams and business strategists. In the pre-processing stage, these actors have an opportunity to shape the course and direction of their target data by defining appropriate variables that would assist most in relationship-finding efforts and predictive outcomes.

### Image: Facebook

Social media sites – such as Facebook - are active users of Data Mining techniques in search of consistent patterns and/or systematic relationships between variables. This does not need much explaining.

- What is the value of this information to a social media site like Facebook?
- Why would they be interested in mining for this information?

Also, worth following the link below if this interests you:

[Click here. \(Links to an external site.\)](#)

---

## Chapter Navigation

Business Intelligence

[7.0 Chapter Overview](#)

[7.1 Analytics and Business Intelligence \(BI\)](#)

[7.2 Data Mining](#)

[7.3 Modelling and Methods](#)

[7.4 Chapter Summary](#)

## 7.3. Modelling and Methods

### Image: Modelling and Methods

#### Method Sophistication

Davenport and Harris (2007) distinguish several sophistication levels of analytics. A higher sophistication level should lead to larger competitive advantage. On a broad level, they distinguish between three areas; access, reporting, and analytics. Access and reporting is mainly descriptive in nature - it is also referred to as market or customer knowledge-focused analysis. The main objectives are to learn more about markets and customers and to provide management with information on marketing and customer metrics, such as market shares, brand awareness, retention rates, and customer profitability. These types of analyses focus on what has happened (past). Usually, simple statistical techniques are used, such as calculating averages.

Analytics involves more sophisticated statistical techniques and answers more complicated business problems. It focuses on the future and is more prescriptive in nature. Analyses may answer questions such as, 'What is the optimal product assortment to offer in a store?' or 'Through which channels should we contact a customer in order to optimise customer value?'. Davenport and Harris' (2007) hypothesis is that firms using these types of analytics should maintain a competitive market advantage (Kooze, Verhoef and Walk, 2016).

#### Methods. Methods Everywhere.

Now that we have explored some of the data-driven approaches to sourcing data and developing an approach for analysis, we can think about methods of analysis. Of course, these methods are purely suited to the type, source, and structure of the data set we are looking at, as well as what we are wanting to achieve with our analyses. Some of the fundamental exploratory analytic methods are:

- Trend analysis
- Geographical (spatial) analysis
- Sampling
- Text mining
- Sentiment analysis
- Web analysis and the many online methods

A more complete guide to analysis methods can be found in many textbooks and online sources. Refer to the extra resources section at the end of this chapter for further resources.

### Image: Trend Analysis

Trend analysis is a fundamental visualisation technique to spot patterns, trends, relationships, and outliers across a time series of data. One of the most basic yet very powerful exploratory analytics, trend analysis can quickly uncover customer, operational, or product trends and events that tend to happen together or happen at some period of regularity. Trend analysis can also yield mathematical models for each of the trend lines. These mathematical models can be used to quantify recurring patterns or behaviors in the data. The most interesting insights from the trend lines can then be flagged for further investigation by the data and informatics team. Firms are frequently interested in how a market develops or how brand sales will grow. For them, this is important; they can constantly monitor market attractiveness and, for planning purposes, aim to know which level of brand sales can realistically be expected. For this purpose, firms might be interested in trends (is the market growing?), but also in actual forecasts. Notably, trends can be used in forecasting models as well.

## Image: Spatial/Geo Analyse

Geographical or spatial analysis includes techniques for analysing geographical activities and conditions using a businesses' topological, geometric, or geographic properties. For example, geographical analysis supports the integration of postcode and government economic and demographic data with a client's internal data to provide insights about the success of a businesses' geographical reach and market penetration. Geographic analysis can be combined with other methods such as trend analysis in order to identify patterns and changes in key markets. It is most useful and widely used to determine the success of marketing and advertising campaigns.

## Sampling

Data sampling is a statistical analysis technique used to select, manipulate and analyse a representative subset of data points in order to identify patterns and trends in the larger data set being examined. Traditionally, sampling methods have been created using a subset of the data called 'sample'. Once the model was created, a second larger set of data was used to test or validate the model. This process was often iterative. Once the model was validated, the model was used to evaluate or 'score' all the data.

With Big Data we are slowly beginning to use more or all of the data to create the model. By doing this, data scientists are able to introduce additional predictor variables into the model in order to increase the lift (accuracy) of the model. When more historical Big Data is used, the model can identify trends that are outside and beyond the time horizon of the truncated cycles that were used in the historical data extraction

technique. MapReduce processes are especially relevant here: we know the ‘reduce’ function makes it easier to work with smaller data sets, and a more compact summary makes data more manageable.

Image: Text and Sentiment Mining

Text mining refers to the process of deriving usable information from text files. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Basically, text mining creates structured data out of unstructured data.

The many sources of text can include:

- Customer Support Channels
- Tech Support and Operations
- Emails and Memos (Staff and Customer)
- Advertising and Marketing Channels
- Human Resources Feedback
- Competitor Response and Contact

We now have processes which allow a more sophisticated ‘reading’ of this text: Sentiment. Sentiment analysis can provide a broad and general overview of customers’ sentiment toward companies and brands. Sentiment analysis can be a powerful way to glean insights about the customers’ feelings about a company, products, and services out of the ever-growing body of social media sites, including Facebook, LinkedIn, Twitter, Instagram, Yelp, and SnapChat.

Web Analytics

Web analytics aim to understand how the consumer uses websites and visits online stores. Web analytics can be viewed as a new Big Data application area of standard descriptive techniques, as it involves new data sources, which are typically more massive than standard data sources. The process gained attention early in 2000 when early online marketers tried to understand how consumers use websites and visit online stores. Consumer online behavior is different from how, for example, customers visit supermarkets; the choice is personalised, dynamic and highly interactive in terms of the context. Some of the most common web analytics methods include:

| Method                    | Description                                                                                                                                                                                            |
|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Clickstream Analysis      | Clickstream data is defined as the electronic record of Internet usage on servers or third-party servers. It is primarily used to analyse behavior on                                                  |
| Customer Journey Analysis | This allows businesses to take data from multiple touch points (television, retail stores etc.) and better understand the journey consumers have made to inform themselves about a product or service. |

| Method               | Description                                                                                                                                                                                                                                                                         |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Attribution Modeling | Development of 'path to purchase' models which look to determine the touchpoints. This aims to assess the effects of each of these on conversion price comparison websites or SEO/SEM).                                                                                             |
| Dynamic Targeting    | A more focused form of market segmentation that groups consumers by behavioural patterns they display when purchasing. This enables insight into behaviours and allows for tailoring of service. These segments can also be referenced with demographic and psychographic criteria. |

### Image: Graph Analysis

Graph analysis is one of the more powerful analysis techniques made popular by social media analysis. Graph analysis can quickly highlight customer or machine (think back to the internet of things) relationships obscured across millions of social and machine interactions. Graph analysis uses mathematical structures to model paired relations between objects. A 'graph' in this context is made up of 'vertices' or 'nodes' and lines called 'edges' that connect them.

### Image: Social Network Analysis

Social network analysis is an example of graph analysis. It is used to investigate social structures and relationships across social networks. It looks also to characterise networked structures in terms of nodes (people or things within the network) and the ties or edges (relationships or interactions) that connect them. While graph analysis is most commonly used to identify clusters of 'friends', uncover group influencers or advocates, and make friend recommendations on social media networks, graph analysis can also look at clustering and strength of relationships across diverse networks such as ATMs, routers, retail outlets, smart devices, websites, and product suppliers. Gephi is a great free software that looks to visualise these connections.

[Read more here.](#) (Links to an external site.)

### Image: Potential Use Cases for Big Data

Customer segmentation is nothing new to an experienced marketer. Traditional B2C segments based on demographic, psychographic and behavioral data are taught in introductory college courses, and B2B marketers are well familiar with segments based on company size or purchase criteria. While these segments have served as a useful guide for decades, the era of big data has created segmentation 2.0. Thanks to a multitude of new data types and improved customer segmentation analysis, we are learning even greater, more granular details about audiences everywhere.

With data collected via social media, mobile and other online interactions, marketers can further narrow these customer segments. These smaller groups, or micro-segments, allow marketers to create more tailored and targeted content, offers, products and services. The more direct and customized an approach is, the higher its potential for real and substantial returns. And while the benefits make sense, it can be difficult to understand how to use big data to further the goals and objectives of your organization. Fortunately, creativity is contagious, and we can learn a lot from the efforts of others. Here are a couple of examples from companies making good use of big data for micro-segmentation.

(Qubole, 2015)

## Time Warner

Time Warner, the massive media company, has millions of customers. Not surprisingly, they collect information on each and every one of them. As a result, they are gathering 0.6 terabytes of data a day. This information has allowed them to create personalised advertising and tailored campaigns for their customers. Time Warner combines demographic data sets and local viewing habits with other data, like real estate records and voter registration. This helps Time Warner understand their customers on a more individualistic level, with insights into income, local environment and political preferences. From there, Time Warner can create and push their campaigns and ads using multiple mediums, like mobile apps, social media, their website and television. This in turn provides more information on how people respond on each of these platforms, and Time Warner can make adjustments and corrections to further customise future ads.

## Amazon

If you have a moment, head over to Amazon.com. What items are they recommending that you buy? They're probably different than what the person next to you would see. But how do they do it? Amazon creates a site unique to each customer. Every time you sign into your account, browse items or make purchases, Amazon takes note. Then, the next time you return, they'll offer you similar items based on your history. Amazon even takes things a step further. By collecting information on all its users, it recognises trends in people who make similar purchases. For example, if more than 80% of people who buy a pair of running shoes also buy a fitness band, you'll be offered a fitness band with your next shoe purchase. By offering people more of what they want, even if they weren't specifically searching for it, they're more likely to make additional purchases.

## Read: Discovery of Topical Authorities in Instagram

An interesting article on the ways modelling, sampling, and web analytics are being used for predictive marketing purposes using Instagram. Author bios are compared to followers and interest groups using some very sophisticated web analytics methods. Note, the article is published currently by Facebook for commercial disclosure.

[Read here. \(Links to an external site.\)](#)

## Watch: Global Leaders Faculty

Ms Claudia Gonzalez Romo, Chief of Public Advocacy for UNICEF, explores consumer data and influencer strategies.

---

## Chapter Navigation

### Business Intelligence

#### [7.0 Chapter Overview](#)

#### [7.1 Analytics and Business Intelligence \(BI\)](#)

#### [7.2 Data Mining](#)

#### [7.3 Modelling and Methods](#)

#### [7.4 Chapter Summary](#)

## 7.4. Chapter Summary

## References

Dell (2015) “Data Mining Techniques” ([Links to an external site.](#)), accessed online at: <http://www.statsoft.com/Textbook/Data-Mining-Techniques> [Feb, 2016]

Saragoza, A. (2013) ‘Why Data Artisans are the new Data Scientists’, Fast Company, [Online] <https://www.fastcompany.com/3020211/why-data-artisans-are-the-new-data-scientists>

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Ohlhorst, F. J. (2012). Big data analytics: turning big data into big money. John Wiley & Sons.

Qubole (2015) “Big Data and Customer Micro-Segmentation: Applications in Media and E-Commerce”, Jonathan Buckley (Ed.), accessed online at: <https://www.qubole.com/blog/big-data/big-data-customer-microsegmentation/> [Feb, 2016]

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Schmarzo, B. (2015). Big Data MBA: Driving Business Strategies with Data Science. John Wiley & Sons.

Teradata (2013) “[\(Links to an external site.\)](#)Listening to 100M Customers with Unified Data Architecture”, accessed online at: <http://blogs.teradata.com/customers/listening-to-100m-customers-with-unified-data-architecture/> [Feb, 2016]

Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge.

Webopedia (2016) “Business Intelligence”, accessed online at: [http://www.webopedia.com/TERM/B/Business\\_Intelligence.html](http://www.webopedia.com/TERM/B/Business_Intelligence.html) [Feb, 2016]

## Exercise

Try to briefly answer these review questions:

1. Explain the difference between Business Intelligence and Data Science. How does the function of Big Data change according to these different approaches (...if at all)?
2. Discuss how the ‘listening cycle’ maximizes data effectiveness through incorporation of the organisation. Describe the key stages that need to occur and why modeling is important.
3. Justify why an online e-tailer would be interested in data mining analytics. What could this illuminate that is easily actionable? How is this information monetised?
4. Describe two of the analytic processes discussed in this chapter and explain why:
  1. This is a unique approach to interpretation, over other approaches.
  2. Some sectors would find this an attractive approach (which ones?).

[Answer here.](#)

## Extra Resources

| Resource                                                                                                    | Description                                                                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                             | Light, easy and fun                                                                                                                                                                                                |
| TED (2013) “What Facebook And Google Are Hiding From The World”, Eli Pariser                                | For those looking to see further information which data is mined on social media sites brings up some interesting perspectives rights and privacy.<br><br><a href="#">Watch here.</a> (Links to an external site.) |
| Aditya Pal, Amaç Herdağdelen, Sourav Chatterji, Sumit Taank, Deepayan Chakrabarti (2016), WWW, Forthcoming. | An interesting article on the ways model web analytics are being used for predict                                                                                                                                  |

| Resource                                                                                                                          | Description                                                                                                                                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                   | <p>purposes using Instagram. Author bios and followers and interest groups using some web analytics methods. Note, the article is currently by Facebook for commercial d</p> <p><a href="#">Read here.</a> (Links to an external site.)</p>                                                                   |
| For the academically curious                                                                                                      |                                                                                                                                                                                                                                                                                                               |
| Barnes, T. J., & Wilson, M. W. (2014). Big data, social physics, and spatial analysis: The early years. Big Data & Society, 1(1). | <p>From the conference comments, the ins paper comes from studying social network importance of addressing privacy issues networks. Development of a model using here is very interesting, with some exce methodology....worth looking at!</p> <p><a href="#">Read here.</a> (Links to an external site.)</p> |
| Proceeding KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.            | <p>An excellent publication-based framework and analytics literature, looking at some around classification, as well as the evolution identification of the Big Data corpus.</p> <p><a href="#">Read here.</a> (Links to an external site.)</p>                                                               |
| 26 popular techniques for analysing Big Data (2014).                                                                              | <p>A more complete guide to analysis methods many textbooks and online sources. We following for a more complete overview</p> <p><a href="#">Read here.</a> (Links to an external site.)</p>                                                                                                                  |
| Data Analytics Methods and Techniques (2014).                                                                                     | <a href="#">Download here.</a> (Links to an external site.)                                                                                                                                                                                                                                                   |
| Beyond the hype: Big data concepts, methods, and analytics (2015).                                                                | <a href="#">Read here.</a> (Links to an external site.)                                                                                                                                                                                                                                                       |

---

## Chapter Navigation

[Business Intelligence](#)

[7.0 Chapter Overview](#)

[7.1 Analytics and Business Intelligence \(BI\)](#)

[7.2 Data Mining](#)

## Chapter 8

### 8.0. Chapter Overview - Visual and Predictive Analytics

This chapter introduces methods of data preparation, information visualisation and prediction. It will introduce predictive analytics, which builds upon our data mining and modelling from the previous chapter. Clustering and association methods, which draw inferences through pattern discovery in order to predict future behaviour and inform future practice, will also be looked at.

This chapter will acquaint you with the different methods that can be used to accelerate the business user and data science team collaboration process, as well as the role that visualisation can play in the interpretation of data through enabling of discussion and comprehension.

1. Data Reduction.
2. Predictive Analysis.
3. Data Visualisation.

---

### Chapter Navigation

#### Visual and Predictive Analytics

[8.0 Chapter Overview](#)

[8.1 Data Reduction](#)

[8.2 Predictive Analysis](#)

[8.3 Data Visualisation](#)

[8.4 Chapter Summary](#)

### 8.1. Data Reduction

What is Data Reduction?

We have been considering Big Data, i.e. data that is becoming bigger and bigger, daily. In this chapter we begin our discussion with data reduction. Why? It helps you simplify the data by removing unnecessary elements, and help you focus on certain variables or the constructs (made from these variables), which will offer some meaning that helps you predict future or have a predictive value. What does that mean?

In simple terms, data reduction is how the empirical digital information (which is getting bigger daily) is transformed into an orderly format. The concept is about reducing the data into meaningful parts. It would involve some sorting, collating, editing, scaling, coding and effectively creating some categorised tables. Consider this example from Astronomy (Source:Wikipedia, 2016). The Kepler Satellite (NASA) records 95 megapixel images once every six seconds, generating tens of megabytes of data per second. In order of magnitude, it is more than the downlink bandwidth of 550 kilobytes per second. The on-board data reduction includes co-adding the raw frames for 30 minutes, reducing the bandwidth by a factor of 300. Furthermore, interesting targets are pre-selected and only the relevant pixels are processed, which is 6% of the total. Such reduced data is then sent to Earth to be processed further. So what does this mean or imply for business managers?

## Why Data Reduction?

Data reduction can help businesses create extra capacity, within the existing environment. Specifically, it helps a business manage continuously expanding data. It may not reduce storage costs but can decrease the amount of required storage capacity in a Storage Area Network (SAN) environment within businesses of any size. Storage Area Network provides access to consolidated, block level data storage. SANs are mainly used to enhance storage devices such as tape libraries and optical jukeboxes, accessible to servers. To the operating system (O/S), the devices would appear as locally attached. It has its own network of storage devices not usually accessible through a local area network by other devices.

We are living in a world of increasing BYO devices that plug into the network and storage needs increase as the data is generated continuously. Therefore, business managers need to strategise for data reduction. There are three strategies that businesses use for data reduction.

## Strategies for data reduction – Pro's and Con's

The first strategy is better known as thin provisioning. This strategy is about eliminating the reserve on unwritten blocks of storage, to make more efficient use of existing storage capacity. With low impact on operations, businesses can save up to 30%. Vendors are available across all sized businesses. However, the existing written data cannot be altered in this strategy to make provision for a larger amount of space.

The second strategy is data deduplication. This strategy identifies repeated data patterns and reduces them to a single instance, so as to help save capacity in the

storage environment. As it reduces repeat patterns to a single physical copy, storage savings can range from 2:1 to 10:1 (based on the industry sector and the data). However, the processes may require more capacity to function accurately. For instance, after the processing, deduplication requires more space to hold the new (processed) data temporarily, before it can be deduplicated. This may impact the amount of actual saved space available in the storage.

The third strategy is compression. It is a strategy that has been tried and tested for almost 25 years. IBM tape drives used this technique two decades ago, by finding repeat patterns of similar information that can be replaced, with a streamline data structure. Space savings can vary dependant on specific types of data. However, this technique introduces latency before and after transfer of data. For example, the files need to be 're-hydrated' or in the original form before it can be accessed, to be compressed. On the other hand, compression algorithms can slow down the writing process.

Recently, a technique called lossless compression – a class of data compression algorithms that allow original data to be reconstructed from compressed data. This has many applications. For example, you may be familiar with FLAC and ALAC file types. These are types of audio files that would contain the same data as a WAV file, which you may be familiar with. WAV takes up much storage, while FLAC and ALAC files use a lossless comparison to create smaller files that take less space.

[Read more here \(Links to an external site.\)](#)

### Watch: Global Leaders Faculty

Prof. Adrian Cheek, Professor of Pervasive Computing for City University London, covers governance, information and technology.

### Data Dimensionality and Reduction

A business may be interested in certain attributes of data such as customers, products, time, stores etc. These are dimensions of the data. For the data warehouse user (or the person who uses data to undertake predictions), these dimensions are entry points to quantitative facts such as sale, profit and revenue that a business wishes to monitor. For example, a business may be interested in knowing a specific jeans that was sold in a store in Melbourne, last month. The data dimension can be hierarchical. For example, to aggregate, summarise and present the data with reduced dimensions, the days can be grouped into months, or quarters can be grouped into calendar years.

### Read: Reducing data dimensions

Reducing data dimensions is usually done via column formats. However, as the number of columns increase, the dimension of the data also rises. This is cumbersome, and

therefore, it dimensionality needs to be reduced. Please read the article to understand seven techniques for reducing data dimensionality.

[Download here. \(Links to an external site.\)](#)

Image: Typical Analysis using Data

As illustrated in this diagram, there are many typical analyses using data. However, we are more focused on predictive analytics and visualisation using the 'reduced and aggregated' data, and the techniques illustrated in the previous sections.

---

## Chapter Navigation

### Visual and Predictive Analytics

[8.0 Chapter Overview](#)

[8.1 Data Reduction](#)

[8.2 Predictive Analysis](#)

[8.3 Data Visualisation](#)

[8.4 Chapter Summary](#)

## 8.2. Predictive Analysis

### What is Predictive Analysis?

The distinction between explanatory and predictive analyses is quite well known and is rather important. Explanatory analyses focus on the 'Why?' question. A researcher aims to know why specific phenomena happen. Predictive Analysis (PA) focus on forecasting marketing and customer metrics. Market forecasting typically focuses on sales forecasting of new products and forecasting market share. At the customer level, this may involve forecasting response to marketing actions, such as direct mails, emails, but also forecasting churn, product returns, and lifetime value. PA can be done rather simply, perhaps by using a kind of weighted average to predict future sales, or in a more sophisticated fashion by using forecast models, such as time-series models or choice models.

We will now look at a brief introduction to two of the most common methods:

- Cluster Analysis
- Association Analysis

## Image: Algorithms and Models

The above analytic algorithms start to move the data scientist beyond the data exploration stage into the more predictive stages of the analysis process. These analytic algorithms are more actionable, allowing the data scientist to quantify cause and effect and provide the foundation to predict what is likely to happen and recommend specific actions to take.

Cluster analysis or clustering is the exercise of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups (clusters). Cluster analysis is used to uncover insights about how customers and/or products cluster into natural groupings in order to drive specific actions or recommendations. For example: Target marketing or maintenance scheduling. Clustering analysis can uncover potential actionable insights across massive data volumes of customer and product transactions and events. It can uncover groups of customers and products that share common behavioral tendencies.

## Image: Association Analysis

Association analysis is a popular algorithm for discovering and quantifying relationships between variables in large databases. Association analysis shows customer or product events or activities that tend to happen together, which makes this type of analysis very actionable. For example, the association rule {eggs, flour} → {baking} found in the point-of-sales data of a supermarket would indicate that if a customer buys eggs and flour together, they are likely to also buy butter. Such information can be used as the basis for making pricing, product placement, promotion, and other marketing decisions. One very actionable data science technique is to cluster the resulting association rules into common groups or segments.

---

## Chapter Navigation

### Visual and Predictive Analytics

#### [8.0 Chapter Overview](#)

#### [8.1 Data Reduction](#)

## 8.3. Data Visualisation

### Visualising - How to Make it Insightful?

Visualising data is a technique to facilitate the identification of patterns in data and presenting data to make it more consumable. Techniques such as charts, graphs, and dashboards have been used for decades to synthesise data into a cohesive and comprehensible format for business analysts, managers, and executives. These techniques have been used to differentiate the contexts of the data that is being visualised in order to:

| Describe                                                                  | Report                                                                                                                                                                      | Observation                                                                           | Discover                                                       |
|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------------|
| Attempting to explain the thing being described, for basic comprehension. | Summarising findings from the past as of a point-in-time. However, as we move into the next stage of visualisation, we move beyond these initial intents into the realm of: | Viewing data to identify significance or patterns which unfold over a period of time. | Interacting to explore, understand, and create meaning between |

Visualisation also has benefits over static data presentation through enabling dialogue: demonstration of other perspectives and relationships. In a business setting, this abstraction can promote discussion, as well as create shared meaning.

However, the sheer volume of data does not always suit visualisation methods: millions of data points, in an image for example, is not always simple to understand and comprehend. New techniques and tools are emerging that utilise exciting new visualisations and animations to visually depict a story about data that far exceeds the standard charts, graphs, and dashboards. Data artisans (Saragoza, 2013) can create these new and dynamic visualisations, with skills that transcend science, design and art (Schmarzo, 2015).

### Visualisation Attributes

Visualisation refers not only to a set of graphical images but also to the iterative process of visual thinking and interaction with the images. An interactive visualisation environment, in which the user may choose to display the data in many different ways, encourages data exploration. One goal of data exploration is the recognition of pattern and the abstraction of structure and meaning from data. As visualisation is becoming

more commonplace, data artisans are using many different dimensions to represent and/or evaluate data. Some of these include:

- Spatial/geospatial: position, direction, velocity
- Temporal, periodicity: state, cycle, phase
- Scale, granularity: weight, size, count
- Relativity, proximity
- Value, priority
- Resources: energy, temperature, matter

A static visual representation can never address multiple dimensions effectively, nor can it effectively show change over time. A series of static representations can only approximate change through periodic snapshots. But looking at and observing data through visualisation, even complex animations, isn't the same as interacting with it to uncover deeper meaning. It requires an effort to traverse and explore the data to uncover these various dimensions. Therefore, they can assist in our ability to explore and engage for discovery, interpretation, and deeper understanding. (Minelli, Chambers & Dhiraj, 2012)

Have a look at how a simple (but highly effective) visualisation of wind patterns across the United States change how we interpret these weather patterns, follow the link below.

[Click here.](#) (Links to an external site.)

Think about how the visualisation is accessible, and how its presentation enables understanding and engagement.

### Visualisation Essentials

When Big Data is visualised there is a common DNA that must be adhered to for this data exploration and comprehension to be encouraged.

All visualisations must ensure:

| They are rhetorical acts                                                                                                                                                                   | They are abstractions                                                                                                                                                                                                | They work on multiple levels                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| This means that they are asking deep, value-focused questions. What really matters here? What does the organisation (and it's customers) really care about? How can we describe our world? | This means we are looking at broad ideas rather than isolated events in our calculations. While idealist and sometimes hypothetical in analysis, they need to be firmly anchored to pragmatic and feasible outcomes. | As you have seen from this section – from visualisation to emotional, spatial and evaluations |

### Read: Visualisation Examples

Dear Data is a project by Stefanie Posavec and Giorgia Lupi. Both tracked everyday things during a week, such as how many times each picked up the phone, and then

visualised the data on a postcard. Then they mailed the postcards to each other, with Lupi currently in New York and Posavec in London. Not commercial at all, but very nifty!

[Read here.](#) (Links to an external site.)

An Introduction to Machine Learning- Stephanie Yee and Tony Chu clarified this tough concept with an amazing scrolling visual example.

[Read here.](#) (Links to an external site.)

Heat Wave - Bloomberg split up the global temperatures and averaged them, added visualisations and voila!

[Read here.](#) (Links to an external site.)

Stereotypes - Developed by the Bocoup Data Visualisation Team, explores the many tropes in films and the the adjective used to describe them. Some are unique to a trope and some words span multiple tropes and genders.

[Read here.](#) (Links to an external site.)

**Image: WHO and GIS**

A geographical information system (GIS) is a computer system for capturing, storing, checking, integrating, manipulating, analysing and displaying data related to positions on the Earth's surface. It is thus a way of linking databases with maps, to display information, perform spatial analyses or develop and apply spatial models.

## **Read: Going from Playtime to Efficiency with Data Visualization**

How to streamline and optimise your visualisations- very handy checklist and with some nice notes on formatting. Invaluable

[Read here.](#) (Links to an external site.)

## **Don't Forget Quality Data**

This course has focused on what we do with the data in order to process and extract for maximum insight. One of the most fundamental elements of this process is ensuring the product is sound: that your data is 'good'!

Data quality consists of several dimensions:

- Completeness of data
- Data being up to date

- No mistakes in data

Completeness of data refers to whether all available data are present for all customers. Mistakes frequently occur, especially with regard to customer descriptors, such as name and address. These mistakes may arise as customers or operators do not complete data entry. The data being up to date concerns whether the data are being updated on a very frequent basis. A database that is not up to date can easily contain mistakes on all kinds of variables, from purchase history to contact details, leading to unreliable information. Mistakes can potentially have strong negative reputational consequences for a firm. For example, a firm may continue sending mail to someone who has recently passed away. Moreover, data being not up to date may also cause wrong predictions to be made on future customer value, which might lead to less than optimal strategies. (Verhoef, Kooge & Walk, 2016)

## The Technology Enabling Analysis

Now that we have covered some of these methods, you must be asking; how do we perform these tasks?; what technology is available? All the traditional players such as SAS, IBM SPSS, Matlab, Tableau, Pentaho, and others are working toward Hadoop-based Big Data analytics. There are also new commercial vendors and open-source projects evolving to address the growing appetite for Big Data analytics.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                   |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| Alpine Data Miner has a cross-platform analytic workbench.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <a href="#">View<br/>extended</a> |
| R is by far the most dominant analytics tool in the Big Data space. R is an open-source statistical language with constructs that make it easy for data scientists to explore and build models. R is also renowned for the plethora of available analytics. There are libraries focused on industry problems (i.e. clinical trials, genetics, finance, and others) as well as general purpose libraries (i.e. econometrics, language processing, optimisation, time series). There is also commercial applications via Revolution Analytics. | <a href="#">View<br/>extended</a> |
| Some of the open-source technologies we have discussed previously, including Apache Mahout, a scalable, Hadoop machine learning library are slowly being used for enterprise function (Schmarzo, 2015).                                                                                                                                                                                                                                                                                                                                      | <a href="#">View<br/>extended</a> |
| RapidMiner provides you with some end-to-end solutions. It helps you prepare data, helps with data reduction, modelling, prediction and visual analysis. Essentially, tools such as these empowers a current business manager.                                                                                                                                                                                                                                                                                                               | <a href="#">View<br/>extended</a> |
| Horizontal Big Data Apps (machine log analytics by Splunk)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <a href="#">View<br/>extended</a> |
| Vertical Big Data apps (telecommunications analytics by Guavus)                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <a href="#">View<br/>extended</a> |

There are also some emergent Big Data apps coming to fruition due to the demands of Big Data volumes and the pace of innovation. Both horizontal Big Data Apps and vertical Big Data apps are emerging. Big Data apps are designed to address specific

business problems and incorporate deeper, more complex prescriptive analytics typically, while also allowing business analysts the ability to explore the data.

### Read: Seven data science lessons from McGraw-Hill Education analytics guru

A very practical interview with Alfred Essa, VP of analytics and R&D at McGraw Hill. This is a very insightful interview which covers how McGraw Hill are applying analytics, but also the role of the data science and an emphasis on predictive methods of data exploration in the future.

[Read here.](#) (Links to an external site.)

## The Actioning Analysis Process

We have looked at the 'why' of analysis, as well as some of the methods to perform the actual analysis. What is missing is the 'how' we make this data actionable! Schmarzo (2015) has illuminated a straightforward process to turn this data into actionable insight and assist in decision making.

1. Establish a hypothesis that you want to test.
2. Identify and quantify the most important metrics or scores to predict a certain business outcome.
3. Employ the predictive metrics to build detailed profiles for each individual customer with respect to the hypothesis to be tested.
4. Compare an individual's recent activities and current state with his or her profile in order to flag unusual behaviors and actions that may be indicative of a customer retention problem.
5. Continue to seek out new data sources and new metrics that may be better predictors of attrition. This is also the part of the data science process to continuously try to improve the accuracy and confidence levels of the metrics and scores using sensitivity analysis and simulations.
6. Integrate the analytic insights, scores, and recommendations into the key operational systems (likely CRM, direct marketing, point of sales, and call center for the customer retention business initiative) in order to ensure that the insights uncovered by the analysis are actionable by frontline or customer-engaging employees.

Once this process is complete, review the findings with the leadership team, and those responsible for responding to the findings (often subject matter experts). Ask them:

- Is the insight of strategic value to what you are trying to accomplish?
- Is the insight actionable (i.e., is the insight at a level upon which I can act on that insight)?
- Is the insight of material value (i.e. is the value of the insight greater than the cost to act on that insight)?

This use case looks at applications of analysis methods examined in this course.

Please read these approaches carefully and think about responding to the questions.

### **Watch: A data analytics strategy for boosting business efficiency and revenue**

A short and simple interview with Michael Koukounas a leading analyst at Equifax looks at the value of analytics which touches on many themes we have covered in this topic: asking the right questions, putting in procedures, as well as some insightful use cases.

[Watch here.](#) (Links to an external site.)

## **Fraud Detection**

Goal: Predict fraudulent cases in credit card transactions.

Approach: Use credit card transaction data and the information on the account holder as attributes.

What are these attributes? What do we want to know about them?

- Label past transactions as fraud or fair. This forms a defining or 'class' attribute.
- Learn a model from how these patterns have formed.
- Use this model to predict future fraud by observing credit card transactions on this account.

What types of methods are used here in this process?

## **Market Segmentation**

Goal: Subdivide a market into distinct subsets of customers where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix.

Approach: Collect different attributes of customers based upon their geographic and lifestyle related information.

What sources would we use here? How could we reliably access this data?

- Find groups of similar customers.
- Measure the quality of these groups by observing buying patterns of customers in the same group versus those from other groups.

What types of methods are used here in this process?

## **Supermarket shelf management/Inventory**

Goal: To identify items that are purchased together by sufficiently many customers.

Is this the only goal here? What other business and commercial benefits can you think of?

Approach: Process point-of-sale data collected with barcode scanners to find dependencies among items.

What can this be cross-referenced with for deeper insight?

- Develop rules and associations between these purchases.

What types of methods are used here in this process?

**Read: Best of Visualisations, 2015**

Looking for more examples of Visualisation? Perfect resource!

[Read here.](#) (Links to an external site.)

---

## Chapter Navigation

### Visual and Predictive Analytics

[8.0 Chapter Overview](#)

[8.1 Data Reduction](#)

[8.2 Predictive Analysis](#)

[8.3 Data Visualisation](#)

[8.4 Chapter Summary](#)

## 8.4. Chapter Summary

In this chapter, we studied the concept of data reduction and dimensionality reduction, to prepare the data for analyses. The field of predictive analytics was introduced which builds upon our data mining and modeling from the previous chapter. Clustering and association methods were looked at which draw inferences through pattern discovery in order to predict future behaviour and inform future practice. This chapter better acquainted you with the different methods that can be used to accelerate the business user and data science team collaboration process. We extended this discussion to the

role that visualisation can play in interpretation of data through enabling of discussion and comprehension. While it is not the expectation of this chapter to turn you into data scientists and number-crunchers, it was our objective to set the foundation that helps you as future business leaders to 'think like a data scientist'. The final actioning your analytics plan provided a pathway for this enablement, and a process by which to make the most of the data you have collected and sourced.

## Exercise

Try answering these brief review questions:

1. What is a key difference between predictive analysis and descriptive analytics?
2. Explain how both cluster and association analysis work to provide future insight and enable market predictions for business leaders. What key insights could they provide to certain industries?
3. Justify why an organisation would invest further to visualise their analysis. What value does this process add?
4. Describe two different dimensions of visualisation representation.
5. Write a sample job posting for a data artisan you are looking to recruit. What skills and requirements should they have and why?
6. Explain the consequences of poor data quality to a business.

Answer here.

## References

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Oehlhorst, F. J. (2012). Big data analytics: turning big data into big money. John Wiley & Sons.

Saragoza, A. (2013) 'Why Data Artisans are the new Data Scientists', Fast Company, [Online] <https://www.fastcompany.com/3020211/why-data-artisans-are-the-new-data-scientists>

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Schmarzo, B. (2015). Big Data MBA: Driving Business Strategies with Data Science. John Wiley & Sons.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge.

## Extra Resources

| Resource                                                                                      | Description                                                                                                                                                                                                                                                                                                                                          |
|-----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                               | Light, easy and fun                                                                                                                                                                                                                                                                                                                                  |
| Koukounas, M. (2013) “A data analytics strategy for boosting business efficiency and revenue” | A short and simple interview with Michael, a leading analyst at Equifax looks at the value of data which touches on many themes we have covered in this chapter: asking the right questions, putting data to work as well as some insightful use cases.<br><br><a href="#">Read here. (Links to an external site.)</a>                               |
| Kirk, A. (2015) “Best of Visualisations, 2015”                                                | Looking for more examples of Visualisation? Here is a great resource!<br><br><a href="#">Read here. (Links to an external site.)</a>                                                                                                                                                                                                                 |
| Techtarget (2016) “Seven data science lessons from McGraw-Hill Education analytics guru”      | A very practical interview with Alfred Esposito, Chief Data and R&D at McGraw Hill. This is a very practical interview which covers how McGraw hill are applying data science, also the role of the data science and analytics in predictive methods of data exploration in business.<br><br><a href="#">Read here. (Links to an external site.)</a> |
| Spotfire (2015) “Going from Playtime to Efficiency with Data Visualization”                   | <a href="#">Read here. (Links to an external site.)</a>                                                                                                                                                                                                                                                                                              |

---

## Chapter Navigation

### Visual and Predictive Analytics

[8.0 Chapter Overview](#)

[8.1 Data Reduction](#)

[8.2 Predictive Analysis](#)

[8.3 Data Visualisation](#)

[8.4 Chapter Summary](#)

---

## Chapter 9

### 9.0. Chapter Overview - Social Media Analytics

This chapter introduces the concept of social media listening/monitoring and examines the process of gathering data from client conversations in the digital media, processing them to obtain insights, leading to information-driven business decisions. Platforms such as Facebook, LinkedIn and techniques of crowdsourcing are discussed.

1. Social Media and Analytics.
2. Crowdsourcing.
3. Digital Marketing.

---

#### Chapter Navigation

##### Social Media Analytics

##### [9.0 Chapter Overview](#)

##### [9.1 Social Media and Analytics](#)

##### [9.2 Crowdsourcing](#)

##### [9.3 Digital Marketing](#)

##### [9.4 Chapter Summary](#)

### 9.1. Social Media and Analytics

Social media is often defined as the future of communication. It consists of countless number of Internet based tools and platforms that enable not only information sharing, but also enhancing its quality. Transfer of text, photos, audio, video and any information is becoming seamless through smart devices such as smartphones and tablets of Internet users.

#### Watch: Global Leaders Faculty

Mr Dion Appel, Managing Director of DDB Melbourne covers how to measure response in the digital economy.

#### Read: The History of Social Networking

The history of social networking - Long before it became the commercialised mass information and entertainment juggernaut it is today, the Internet – and its predecessors – were a focal point for social interactivity. This article takes you through this evolution.

[Read here.](#) (Links to an external site.)

## Platforms

Platforms such as Facebook and LinkedIn have enabled the creation of online communities. People share information as they desire, resulting in content (data) that can be created, shared, searched and be used of analytics. Tools and websites such as Digg and Reddit find specific information search simple by tagging (or assigning) sites searchable with keywords. Twitter is a Micro-blogging site that enables aggregated information possible using hashtags (#). Whatsapp is an instant messenger system that currently enables video and audio exchanges.

## Tools

Social Media Analytics (or SMA) is the practice of aggregating data from blogs and social media websites and analysing this data to inform business decisions. Digging customer sentiment is the most common version to support marketing and customer service. This process stretches beyond monitoring data, or undertaking a basic analysis of retweets or 'likes'. It is the process of developing an in-depth idea regarding a customer, rather at an emotional level, so that it can be used for targeted advertising.

Here are 25 top social media analytic tools:

[Click here](#) (Links to an external site.)

## Watch: Facebook

As per the latest Sensis Report, Facebook remains the topmost preference for social media platforms. What do they do? Watch the video above.

Facebook was founded in 2004 with a mission to connect people across the world. Starting with friends and families, the platform has now evolved to discover the latest happenings in the world, to share and express what matters to connected communities.

An opinion on Facebook does matter, as analytics is able to aggregate consumer opinion based on 'Likes'. The most important feature is that the power is in the hands of the consumer, who can make or break a product or service. The platform has evolved to become the topmost influencer across all age groups. The latest on Facebook is a feature known as 'Instant Articles', a new way to create interactive articles, faster. Watch the video above.

## Read: LinkedIn

LinkedIn enables people to create a professional profile of themselves, and then connect with professionals in the network. The network is not a 'social network' but is rather a professional network. LinkedIn Influencers are selected by invitation only from a pool of 500+ of the world's foremost thinkers, leaders, and innovators. As thought leaders in their geographic locations and industry sectors, they discuss and write articles on trending topics. These may be the future of higher education, the workplace culture at Amazon, the plunge in oil prices, and the missteps of policymakers – just to name a few.

[Read more here.](#) (Links to an external site.)

## Watch: What Facebook And Google Are Hiding From The World

For those looking to see further information into the ways in which data is mined on social media sites, this TED talk brings up some interesting perspectives to online content, rights and privacy.

---

### Chapter Navigation

#### Social Media Analytics

##### [9.0 Chapter Overview](#)

##### [9.1 Social Media and Analytics](#)

##### [9.2 Crowdsourcing](#)

##### [9.3 Digital Marketing](#)

##### [9.4 Chapter Summary](#)

## 9.2. Crowdsourcing

The use of collected intelligence, often gathered from the public, to inform business related tasks is better known as the technique of Crowdsourcing. These tasks are usually done by a third party service provider or another company. Businesses gain insight into their customers and their desires, often free. There is no longer the need to hire full-time staff to gain these insights, as the Internet has given birth to the largest pool of workers – the Crowd!

Below is an example, iStockPhoto.

[View here.](#) (Links to an external site.)

It allows both amateur and professional photographers, and videographers as well as illustrators to upload their work. If the images are bought, they earn royalties. There is often criticism in that this technique kills the professionals. However, in the digital era, this effort is becoming mainstream rapidly.

In 2014, Ebola in West Africa killed approx 11,000 people. At the time, public engagement was required to keep this disease at bay globally and prevent its spread. Enter Healthmap! Healthmap is an online monitoring system developed in 2006 at Boston University, with epidemiologists, researchers and software developers. It has the following features:

- 'healthmap.org' and mobile app 'Outbreaks Near Me' deliver real-time intelligence on a broad range of emerging infectious diseases for a diverse audience including libraries, local health departments, governments, and international travelers.
- It brings together disparate data sources, including online news aggregators, eyewitness reports, expert-curated discussions and validated official reports, to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health.
- Through an automated process, updating 24 hours, 7 days and 365 days a year, the system monitors, organises, integrates, filters, visualizes and disseminates online information about emerging diseases in nine languages, facilitating early detection of global public health threats.

Two years from Ebola, we have Zika virus and there are personal risk profiles being made available in addition to monitoring using mobile phones. Read more and watch the video below:

[Click here](#) (Links to an external site.)

## Watch: In Australia

According to Matthieu Jacomin, director for Australia and NZ at crowdsourcing agency eYeka, local companies have swiftly embraced crowdsourcing as a new way to innovate and market their products. Pop Tops launched a poster campaign. Australia Post launched a social video campaign. Nescafé challenged the eYeka community: how do you reinvent instant coffee to match the quality you get in a coffee shop while keeping the convenience of 'instant', at home?

## Smartphones and Tablets

By end of 2016, smartphones had penetrated every economy, small or large scale. Searching for information is becoming a habit, better known as 'Googling' named after the 'Google' search engine! Sensis report (2016) suggests that in Australia, people owned three Internet-enabled devices, but a rising number is using Smartphones (76%). Males have embraced smartphones, while females have taken to tablets and wearable devices (such as those that monitor heart etc). Smartphones have become almost

universal for under 40's, becoming more common amongst 40-50 age group. Typically, the generation from 60s onwards is gravitating towards tablets.

[Read more here.](#) (Links to an external site.)

## Smartphones, why?

There is an application for practically every service. Let us look at a few examples.

|                     |                                                                                                             |
|---------------------|-------------------------------------------------------------------------------------------------------------|
| Uber                | The ability to track and trace a cab from when you request, to the minute, with driver from a mobile phone. |
| Menulog             | An application that enables you to order food at the convenience of your home delivery.                     |
| Gomoney             | A bank application that enables you to transfer money to any account, using                                 |
| Express Mobile apps | A variety of services accessible via MyGov (including medicare) for Australia                               |

There is practically an application for every person, be it transport, medical care, emergency, banking. Applications have enabled smartphones. As we have seen in chapter 1, the trend towards self-service analytics is increasingly high amongst world population. Think about it – how often have you searched for the nearest café or restaurant? How about checking or using a mobile bank account using a mobile phone?

[Read more here.](#) (Links to an external site.)

## Watch: Global Leaders Faculty

Ms Tracey Fellows, CEO of REA Group explores digital disruption on mobility and data.

---

## Chapter Navigation

### Social Media Analytics

#### [9.0 Chapter Overview](#)

#### [9.1 Social Media and Analytics](#)

#### [9.2 Crowdsourcing](#)

#### [9.3 Digital Marketing](#)

#### [9.4 Chapter Summary](#)

## 9.3. Digital Marketing

### Advertising on Social Media

Australian views (Sensis Report, 2016) are polarised towards businesses advertising on social media. The report shows that 27% of the people have no issues with sponsored posts from businesses they follow (or have opted in), 34% is happy to see ads, and 34% will click on the ads occasionally. Interestingly, 41% indicated that they will inspect any brand's social media presence, before making an online purchase (if they did not use the site previously). It shows that social media based advertising is still an important tool. The report also shows that 31% of social media users are rating products and services of any business they use. According to the report, the rating has begun to dip drastically from 2015. And people are providing ratings in 2-3 months. Putting it all together, advertising on social media is no longer an added optional item for business - it is a requirement to stay in business.

#### Image: Business in Social Media

In the last year, the uptake of social media by businesses has risen in Australia. 48% of SMEs now have at least a social media presence. 79% of large businesses have a social media presence. The average Australian consumer using social media is 69%, although they are not interacting only with businesses. With the increase in the number of applications and optimisation of services on them, businesses are trying to reach out to average customers in their day-to-day life. While smartphones are here to stay, the applications that are provided by every business that touches the daily lives of people are now a requirement. Specifically, it is a necessary evil for the survival of every business. That brings us to the business use of social media analytics – mainly in digital marketing.

### Digital Marketing Trends

According to Smartinsights (2016), the digital marketing landscape is changing rapidly. In this list:

|                                       |                                                                             |
|---------------------------------------|-----------------------------------------------------------------------------|
| Big Data                              | Includes market and customer insight and predictive analytics               |
| Content marketing<br>Communities      | Branded niche or vertical communities                                       |
| Conversion rate<br>optimisation (CRO) | Improving website experiences                                               |
| Display                               | Banners on publishers, ad networks social media including retargeting and p |
| Internet of Things (IoT)              | Marketing applications                                                      |
| Marketing Automation                  | Incl CRM, behavioural Email marketing and web personalisation               |

|                            |                                                        |
|----------------------------|--------------------------------------------------------|
| Mobile marketing           | Mobile advertising, site development and apps          |
| Paid search marketing      | e.g. Google AdWords Pay Per Click                      |
| Online PR                  | Including influencer outreach                          |
| Partnerships               | Including affiliate and co-marketing                   |
| Search Engine Optimisation | SEO or organic search                                  |
| Social media marketing     | including Social CRM and Social Customer Care          |
| Wearables                  | e.g. Apple Watch, activity trackers, augmented reality |

[Read more here.](#) (Links to an external site.)

## SEO

Search engine optimisation (SEO) is a set of strategies to bring in more visitors to a website. This is achieved by getting a high ranking placement in search engine pages such as Google, Bing, Yahoo etc. A website that has been optimised, but yet focusing on relevance and organic ranking is known to be following White Hat SEO practice. The technique of using keywords and keyword analysis, backlinking, link building to link popularity, and writing content are white hat practices. Businesses who intend to make long term investments on their websites, better known as Ethical SEO, are following White Hat practice.

## Pinterest

Pinterest is a platform that works like a Scrapbooking concept, rather a catalogue of ideas. It enables innovation, entrepreneurship by making people think! Read this article below and think about it!

[Click here.](#) (Links to an external site.)

## Read: How Australian people and businesses are using social media

The social media landscape in Australia continues to evolve rapidly for both consumers and businesses. This report for 2016 looks at the impact of social media and business.

[Download here.](#) (Links to an external site.)

## Watch: Global Leaders Faculty

Mr John Fraser, Secretary of the Department of Treasury, Australia, discusses the importance of judgement in national forecasting.

## Question Time

Try answering these brief review questions:

1. Can you name six mobile applications that are used regularly? What are they used for?
2. What is the most popular social media platform in Australia? Why do you think it is popular?
3. What is social media analytics? Why do businesses undertake it?
4. What is crowdsourcing? What is its relevance in digital marketing?

[Answer here.](#)

---

## Chapter Navigation

### Social Media Analytics

#### [9.0 Chapter Overview](#)

#### [9.1 Social Media and Analytics](#)

#### [9.2 Crowdsourcing](#)

#### [9.3 Digital Marketing](#)

#### [9.4 Chapter Summary](#)

## 9.4. Chapter Summary

The objective of this chapter was to give you a taste for social media and its relevance in digital marketing. Different platforms such as Facebook, LinkedIn and Pinterest use different techniques for assisting people and businesses. The concept of Ethical SEO is discussed. We looked at the concept of Crowdsourcing and GIS mapping where consumers are engaged in promoting their own work, connecting with others and obtaining a personalised risk profiles. What does it all mean? Social media analytics is becoming user-driven, and in real time. It is no longer pushed by businesses, it is rather created and run by users themselves. It affects consumers and businesses alike. While consumers look for real answers in the market, businesses require digital marketing techniques to stay in business.

## References

Dell (2015) "Data Mining Techniques" ([Links to an external site.](#)), accessed online at: <http://www.statsoft.com/Textbook/Data-Mining-Techniques> [Feb, 2016]

Saragoza, A. (2013) 'Why Data Artisans are the new Data Scientists', Fast Company, [Online] <https://www.fastcompany.com/3020211/why-data-artisans-are-the-new-data-scientists>

Minelli, M., Chambers, M., & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. John Wiley & Sons.

Ohlhorst, F. J. (2012). Big data analytics: turning big data into big money. John Wiley & Sons.

Qubole (2015) "Big Data and Customer Micro-Segmentation: Applications in Media and E-Commerce", Jonathan Buckley (Ed.), accessed online at: <https://www.qubole.com/blog/big-data/big-data-customer-microsegmentation/> [Feb, 2016]

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. John Wiley & Sons.

Schmarzo, B. (2015). Big Data MBA: Driving Business Strategies with Data Science. John Wiley & Sons.

Teradata (2013) "[\(Links to an external site.\)](#)Listening to 100M Customers with Unified Data Architecture", accessed online at: <http://blogs.teradata.com/customers/listening-to-100m-customers-with-unified-data-architecture/> [Feb, 2016]

Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge.

Webopedia (2016) "Business Intelligence", accessed online at: [http://www.webopedia.com/TERM/B/Business\\_Intelligence.html](http://www.webopedia.com/TERM/B/Business_Intelligence.html) [Feb, 2016]

<http://www.digitaltrends.com/features/the-history-of-social-networking/>

<https://medtechboston.medstro.com/blog/2016/08/16/healthmap-develops-live-zika-tracker/>

<https://www.thebalance.com/what-is-crowdsourcing-marketing-and-how-is-it-used-2295467>

## Extra Resources

|                                                                     |                                                             |
|---------------------------------------------------------------------|-------------------------------------------------------------|
| Timeline for social media                                           | <a href="#">Read here.</a> (Links to an external site.)     |
| Sensis report for Social media in Australia                         | <a href="#">Download here.</a> (Links to an external site.) |
| Branding in the age of Social Media (2016) Harvard Business Review  | <a href="#">Read here.</a> (Links to an external site.)     |
| What Cancer Researchers can learn from direct-to-consumer companies | <a href="#">Read here.</a> (Links to an external site.)     |

## Chapter Navigation

### Social Media Analytics

#### [9.0 Chapter Overview](#)

#### [9.1 Social Media and Analytics](#)

##### [9.2 Crowdsourcing](#)

##### [9.3 Digital Marketing](#)

##### [9.4 Chapter Summary](#)