



THINK BIG:
BRITAIN'S DATA
OPPORTUNITY

◆◆ wandisco

DATA IS THE NEW OIL

CLIVE HUMBY

CHIEF DATA SCIENTIST, STARCOUNT
INVENTOR TESCO CLUBCARD

4	FOREWORD
8	CHAPTER 01 THE DATA EXPLOSION
14	CHAPTER 02 THE OPPORTUNITY WITHIN
20	CHAPTER 03 LAYING THE FOUNDATIONS
24	CHAPTER 04 THE FUTURE OF DATA
30	CONCLUSION
32	PREDICTIONS & RECOMMENDATIONS
34	ABOUT US



➤ FOREWORD

Data rules the world. This isn't just what they say out in Silicon Valley – it's a simple matter of fact.

We create more data every 20 minutes than is currently held by the Library of Congress. More than 90 per cent of all the data in the world was generated in the last two years alone. Today, every company – from Aston Martin to Zoopla – is to a greater or lesser extent a data company that churns out colossal levels of information by the second.

This has led to the rise of what has become known as 'Big Data': information sets so large and complex that to understand them using traditional methods would be impossible. An increasing number of organisations hope that using new technology to store, query and analyse this data in real time will help them better understand their customers' behaviour.

Information that was previously thrown away is transforming every facet of human existence. And what was once dismissed as useless is essential to devising data-driven solutions to universal problems, offering the potential to make us safer, happier and healthier. We shouldn't think about Big Data as simply being an endless bundle of information. Instead we should think of it as providing the opportunity of competitive advantage across every industry.

A study by McKinsey estimated that data-driven strategies could save up to \$100 billion annually across the US healthcare system alone, helping to optimise innovation, improve research and ultimately build new life-saving tools for medical staff. SAS, a software provider, predicts the UK government could save up to £2 billion by implementing Big Data solutions to detect fraud¹.

The UK government has named Big Data as one of 'eight great technologies' for the future, predicting that it will generate £216 billion and provide 58,000 jobs for the economy by 2017. Almost £50 million has been set aside to fund The Alan Turing Institute and provide a British base for Big Data and algorithm research.

Despite this momentum, Big Data has its critics. In March 2014, *FT Weekend* published a four-page feature with the title 'Big Data: are we making a big mistake?'; while in June the same year *The Guardian* ran with 'Big Data: saviour or sham?'

The authors suggested that Big Data's potential has been overhyped; analysts stand accused of making leaps of logic when interpreting the data, identifying causation where there is none, and making recommendations based on spurious correlations.

While there may be some truth in this, such criticism does not diminish the increasing value of Big Data in various fields, from business to healthcare, from sport to space exploration. The backlash is a typical response to ambitious claims about potential, even if the technology does end up transforming the world. *The Economist* perhaps said it best: "It happened with the internet, and television, radio, motion pictures and the telegraph before it. Now it is simply Big Data's turn to face the grumblers."

The truth is that with more information at our fingertips we stand to learn more about the world around us – and act with greater precision, speed and efficiency. In *Think Big: Britain's Data Opportunity*, we outline the Big Data opportunity to show how organisations big and small are capitalising on its potential.

Companies of the future will live or die by their analytical skills and the Chief Data Officer could well be the future king of the boardroom. The business heroes of the coming years may be more scientific than artistic, able to make insightful commercial conclusions based on empirical evidence.

Here at WANdisco we believe in the power of Big Data, but our view is the UK isn't yet doing enough to harness the opportunity. What's lacking is an approach that will ensure we have enough data scientists, as well as enough managers and analysts with the know-how to make effective decisions. Our education system is not producing the graduates with the skills to get Britain ready for its data destiny.

It's our belief that every conceivable area of life where things need to be organised and decisions made will transform to become leaner, more efficient, more powerful and more lasting. Welcome to a new era: the age of Big Data.

DAVID RICHARDS
CEO, PRESIDENT & CO-FOUNDER,
WANDISCO

¹ Data equity: Unlocking the value of big data, SAS & CEBR, 2012.

BIG DATA BY NUMBERS



BY 2020 THE AVERAGE BUSINESS WILL HAVE TO MANAGE 50 TIMES MORE INFORMATION THAN IT DOES TODAY

ECONOMIST INTELLIGENCE UNIT
& CAPGEMINI 2012



THERE WILL BE OVER 40 TRILLION GIGABYTES ON EARTH BY 2020 – THAT'S 5,200 GIGABYTES FOR EVERY PERSON

IDC & EMC 2012



BY 2017 THERE WILL BE A TOTAL OF 19 BILLION INTERNET CONNECTED DEVICES IN THE WORLD

WEF 2012



IN 2014 THERE WILL BE AN EXPECTED 77 BILLION APPLICATION DOWNLOADS, COMPARED TO 10 BILLION IN 2010

CISCO 2013



64% OF ENTERPRISES ARE PLANNING BIG DATA PROJECTS IN THE COMING YEAR

GARTNER 2013



THE MARKET FOR BIG DATA TECHNOLOGY AND SERVICES WILL REACH \$16.9 BILLION BY 2015

IDC 2012



OVER THE LAST 10 YEARS THE DIGITAL SHARE OF THE WORLD'S STORED INFORMATION HAS INCREASED FROM 25% TO OVER 98%

IBM 2013



BIG DATA JOBS ARE FORECAST TO INCREASE 92% BY 2017

SAS 2013

CHAPTER 01

01 THE DATA EXPLOSION

- We live in a world dominated by data. Every shopping trip, search queried on Google, Uber-cab hailed and change in the traffic-lights is in some way a data transaction.

It's not necessarily true that there's more data out there than before; rather that, instead of disappearing into the ether, details of how we conduct our daily lives are increasingly being captured, stored and used (in most cases) to make our lives better.

In 2012, the International Data Corporation (IDC) calculated that there was a total of 1.8 zettabytes of data stored in one shape or another across the world, nearly 50% more than in 2010. It said that the global stock of data was therefore doubling, at least, every two years.

For context, an average email might be 50 kilobytes (50,000 bytes) in size – attach a picture and it becomes five megabytes. A large computer hard drive might be able to store a terabyte – or 1,000 bytes cubed – of data, while even an average phone contains 16 gigabytes.

The main reason for this data explosion is the rise of connected devices and software capable of creating files of text, images and video. At the end of 2013, one in every five people in the world had a smartphone, while one in 17 owned a tablet, according to research by *Business Insider*.

The phenomenon has also been fuelled because web users are increasingly comfortable with sharing personal details about themselves when they interact with websites.

Meanwhile, even more data is created by mobile phones automatically through global positioning, Bluetooth, Wifi and other kinds of mobile data that phones produce without users necessarily knowing about it.

In 2012, technology analysts at Gartner concluded that between 10% and 15% of organisations were making proper use of Big Data and added that those that did would see a 20% increase in revenue as a result.

Google, Facebook and Amazon are the giants of the internet age, but above all they are data companies. Demography, buying trends, times, dates, where people come from and where they go are all logged and crunched in a bid to understand users better, give them more of what they want, and make them easier to target for marketers.

- 235 TERABYTES OF DATA SITS IN THE US LIBRARY OF CONGRESS – FACEBOOK ALONE STORES, ACCESSES AND ANALYSES OVER 30,000 TERABYTES.
MCKINSEY
- GOOGLE PROCESSES MORE THAN 24,000 TERABYTES OF DATA A DAY.
ACM – COMPUTING SOCIETY
- THE BIG DATA MARKET IS FORECAST TO BE WORTH \$32.1 BILLION IN 2015 – RISING BY 49.5% TO \$48 BILLION IN 2016 AND 66% TO \$53.4 BILLION BY 2017.
WIKIBON
- THE MOST ADVANCED FLOPPY-DISKS STORED 200 MEGABYTES OF DATA – THE AVERAGE HOME COMPUTER STORES MORE THAN 5,200 TIMES THIS AMOUNT.

Advertisers were some of the earliest adopters of the Big Data approach. Rather than hit and hope, they could target customers who wanted to buy from them. No longer, in theory, would retirees see ads for 18-to-30 holidays nor would students be forced to sit through details of generous new pension plans.

But there are thousands of other applications of Big Data that are at various stages of development. Some are being used widely right now; some are evolving into usable products and yet more are mere concepts of how things could look in the future. Take GRITIT, a UK-based firm that helps to clear roads during periods of heavy snow. It uses Met Office data to organise and allocate jobs to regional digger drivers and log work carried out.

Sending jobs to gritters via smartphone, the company tracks their progress while creating reports in real time. By analysing the many variables in the data, the management team can gauge the success of the job and teams can investigate if something goes wrong – even on nights when there are thousands of site visits.

With every year that passes more and more ideas transfer from the realm of science fiction to that of science fact.

For businesses, Big Data arguably presents the biggest opportunity since the birth of the internet itself. Big Data is not just for sophisticated technology firms and it is being put to use by all organisations in all industries.

Everyone can learn about the world they inhabit in microscopic detail and use the information to their advantage. From pharmaceuticals to sports science, businesses that utilise Big Data will create competitive advantage and open up new market opportunities.

Big Data will not only help businesses make smarter decisions, it will also create a new line of smart products and services that use information to perform at their peak while cutting out waste. One thing is for sure, the data explosion is only going to get bigger.

➤ INTRODUCING BIG DATA

HADOOP

Apache Hadoop is the enterprise framework of Big Data that is used both to store and process incredibly large data sets.

The underlying technology was originally invented by Google to index all the rich textual and structural information they were collating, and present results to users in a meaningful way.

Yahoo developed Apache Hadoop as an enterprise platform incorporating this technology, allowing companies to run extensive analytics of both structured and unstructured data – information that doesn't fit nicely into tables. In the same way Google indexed users' search behaviour, Hadoop lets organisations learn more about customers and consumers.

The system is designed to run on a large number of machines that don't share any memory or disks. When an organisation loads its data into Hadoop, the software separates that information into pieces that it then spreads across different servers – meaning there is no one place where you manage all of your data. Because there are multiple copy stores, data stored on a server that goes offline or dies can be automatically replicated from a known good copy.

CLOUDERA

Cloudera was founded in 2008 to provide the first enterprise-ready implementation of Apache Hadoop, set up by three engineers from Google, Yahoo and Facebook – three businesses that were among the trailblazers of Hadoop.

Based in Palo Alto, California, Cloudera has enjoyed rapid growth since first receiving some \$5 million worth of funding in 2009. The company has since expanded to become one of the leading providers of Big Data solutions, used by a diverse range of companies and organisations including Expedia, BT, Western Union, Nokia and eBay.

In March 2014, the firm received a \$740 million investment from Intel, with its 15% stake valuing the firm at \$4.1 billion. Cloudera is one of a handful of Silicon Valley start-ups to receive multi-billion dollar valuations from investors pre IPO.

By replacing its own Big Data project with Cloudera's solutions, Intel provided a strong indication that Hadoop is the future of Big Data.

HORTONWORKS

Hortonworks does one thing: building, managing and implementing Hadoop. The company has devoted itself to working within the open source space, reaching out to customers through the existing products of its partners such as Microsoft, Teradata and SAP. It houses the largest collection of Hadoop "committers" – the name given to those who add code to the Hadoop mainframe.

Founded in June 2011, the firm's dedication to open source has seen it quickly established as one of the major players within the Hadoop community.

The company was set up by 24 engineers who were involved in the original development of Hadoop at Yahoo, and backed by Benchmark Capital the firm raised \$25 million in November of that year.

Named after Horton the Elephant of the *Horton Hears a Who!* book, the company was described by Forrester Research as the "technology leader and ecosystem builder of the entire Hadoop industry".



➤ **EXPERT VIEW**

CLIVE HUMBY

CHIEF DATA SCIENTIST, STARCOUNT
INVENTOR TESCO CLUBCARD

Clive Humby is chief data scientist at Starcount, the fan science company, and one half of the husband-and-wife duo behind dunnhumby, the firm credited with the invention of the Tesco Clubcard – one of the first instances of Big Data's early success.

I once said that what we are witnessing with Big Data is a revolution reminiscent of that seen with the discovery of crude oil. That remains true today, as we continue to go through a process of sourcing and refining data, to obtain its true value.

As oil became more plentiful, we developed technology to turn it into plastics, food and other powerful applications and as data is today becoming more plentiful, we're getting much better at using it in new and innovative ways.

Big Data, just like crude oil, has the potential to become a revolutionary powerhouse of not only industry but society as well.

Yet its value will only be found in a careful application of analysis and refinement.

At my former company, dunnhumby, we first learned to understand customers by analysing what they were buying, where they were shopping and when – establishing the Tesco Clubcard and transforming Tesco's business model. Today, this is an established process and works well for many retailers.

As our experience and ability to analyse data improves, it is clear that the true benefits of data will come with the intricate mapping

and merging of data sets. This is Big Data, it is different to lots of data you find in a single source such as phone records.

Taking the retail example further, if we now compare what people buy with their TV viewing habits or changes in the weather, we can provide unique and highly specific insights which are already improving the way that businesses serve their customers and how efficiently they run.

Business processes, HR and the customer will all enjoy a massive payback from this revolution.

And this isn't just beneficial for business. Big Data has the potential to be the number one game-changer in healthcare provision globally. From identifying patterns in behaviour and illness, to pre-empting reactions to specific medicines, data has the potential to save millions of lives. The data already exists – we just need to combine it and analyse it.

As we become increasingly reliant on data as a major driver of economic and social development, it is vital we treat it with the careful analysis and application it requires. We must also nurture it. Consumer privacy and fair use of data become important principles if we want the well to keep flowing.

Downtime is no longer an option and the effect of poor analysis will become increasingly damaging. We must ensure our data is safe and that the next generation are fully prepared with the skills required to transform the new crude oil into the enormously valuable asset it has the potential to be.

"BIG DATA HAS THE POTENTIAL TO BE THE NUMBER ONE GAME-CHANGER"

CHAPTER 02

02 THE OPPORTUNITY WITHIN

 THE FUTURE IS NOW

Big Data is often talked about as an asset of the future, but for many organisations it represents the here and now. Businesses across the world are using the information at their fingertips to work smarter by creating products and services that are more relevant to their customer base.

Retailers were among the early adopters of Big Data analytics, perhaps most notably Tesco, whose Clubcard provided the template for others to emulate. Back in the early 1990s Tesco was lagging behind both Marks & Spencer and Sainsbury's. Fast-forward a few years and it was the UK's No.1 retailer.

The story of Tesco's rise to dominance is the stuff of legend, but it was rooted in the premise that if any business is to be successful it has to listen. And as former CEO Sir Terry Leahy said: "The best place to find the truth is to listen to your customer."

By analysing the data gleaned from the company's flagship loyalty card, Tesco was able to extend tailor-made offers to their customers with the aim of increasing the return rate. It's a scheme that has since been imitated the world over, but it is Tesco's Clubcard that continues to set the gold standard.

MET OFFICE

The Met Office receives about 100 million 'observation messages' every day (2014), using an IBM supercomputer that can do more than 100 trillion calculations per second and produce over 20 terabytes of data every day.

TESCO CLUBCARD

Clubcard now has over 43 million holders globally, processing 6 million transactions a day (2009) and the value of points redeemed was £780 million in 2010/11.

Data of the sort mined by Tesco is part of a much broader set that has sparked particular interest – the "found data" that includes the digital exhaust of web searches, credit card payments and mobiles pinging the nearest phone mast. While found data is cheap to collect, it is in essence a messy collage of datapoints, each collected for disparate purposes that is constantly being updated in real time.

Harnessing found data was one of the early challenges encountered by data scientists, with a number of critics pointing the finger at misguided conclusions as reason to be sceptical of data's potential.

The UK Meteorological Office is making the weather a Big Data success story, developing a revolutionary forecasting service. Events in space, such as solar flares and solar wind, can impact the performance of the electricity grid, satellites, GPS systems, aviation and mobile communications used by satellite operators, electricity and aviation industries – even the armed forces.

The project enables the Met Office to analyse large amounts of different types of data, including solar flare imagery from NASA, and provide warnings of space weather events so that the government and businesses can take appropriate action to minimise its impacts. The Met Office aims to deliver public space weather forecasts, providing stakeholders and interested parties with access to real-time space weather information and predictions.

THE DATA DISRUPTORS

Timesaving strategies are becoming increasingly common, helping institutions make the shift to digital while increasing the level of service afforded to customers.

For more than 150 years Western Union has helped make the world a small place, originally delivering telegrams and now processing money transfers that stretch from one corner to the globe to the other.

Today it is one of the world's largest financial services companies, one that processes \$79 billion and generates more than 200 terabytes of data annually.

The firm is applying Big Data analytics to eliminate wire transfer fraud, speeding up a process that takes a fraction of the time demanded by traditional methods. Every day the company taps massive databases to sort through customer information and calculate the risk that a particular transaction might be the result of a scam. When they find a risky transaction, they block it.

To protect consumers Western Union analyses huge streams of data for anomalies, using Hadoop to deal with the complexity of its data queries. Not only is this a lot less expensive and more capable than past data warehouse solutions, the fraud prevention rate is higher and more comprehensive than ever before.

The underlying inspiration behind Big Data applications is competition – the endeavour to gain an advantage over your rival. Perhaps nowhere is this better demonstrated than in the cutthroat sport of Formula One racing.

McLaren's cars are fitted with more than 100 sensors that record thousands of different readings about race conditions, as well as the car and driver's performance. It is said that by the end of an average race enough data has been logged to fill several telephone books.

The sensors are not only deployed across several different cars on race day, but also in testing, simulations and practice laps. The data is interpreted to find out whether performance is optimal or if more tweaking is required. It is compared to previous races and different drivers to get the best possible picture.

But this technology developed for F1 has been revolutionary elsewhere. McLaren Electronic Systems is heavily involved in advanced monitoring for healthcare applications. The tech that was developed to monitor human physiology on the racetrack is the underlying framework for what is used by doctors in hospitals. And at McLaren Applied Technologies, the company has been applying F1-tested manufacturing technology to new products that would benefit from similar structural techniques – such as aerodynamics.

McLaren CIO Stuart Birrell says that Big Data is underpinning everything the company is working on. The firm says that during racing season the car is modified every 20 minutes based on new information gleaned from analytics. But the insight provided by Big Data is only so good as the queries made by its engineers. The challenge, as Birrell says, is to "give these guys who are naturally inquisitive the ability to ask questions that we in IT would have never predicted."

If anything this shows that Big Data does not erase the need for vision or human insight. On the contrary, we must have business leaders who can spot a great opportunity, understand how a market is developing, think creatively and propose truly novel offerings. The successful companies of the next decade will be the ones whose leaders can do all that while changing the way their organisations make many decisions.

➤ EXPERT VIEW

BARBARA HOLZAPFEL
FORMER MD, SAP LABS NORTH AMERICA

Barbara Holzapfel is former head of SAP Labs, the world-leading provider of enterprise software and software-related services.

Big Data is a huge driving force in the wider context of the Internet of Things. Companies are now collecting, aggregating and analysing data at a much higher speed than previously possible. It has opened up a completely new dynamic in how businesses are run and how they interact with their ecosystem.

But what we are seeing now is just the tip of the iceberg and innovation will accelerate at an unprecedented pace over the next few years.

The value proposition is important and it's up to companies like SAP to be clear on what the user benefits are: what people can do now that they couldn't before and how valuable that is to them.

Data is already having massive implications for how people work and live and how business is conducted. It will create new opportunities in the development of new applications and services; anything from network infrastructure, to databases, to the carrier capacity, to the software tools on top of it all. All the existing models for those will be rethought.

To illustrate one such opportunity we could look at a case in San Francisco, where 30 per cent of the traffic in the city comes from drivers trying to find parking spaces. That has a major impact on the local and global environment.

Through innovations in Big Data, you can now find and reserve a clear parking spot near your destination. You reserve and pay for this through your mobile. The benefit to a local coffee company is a tie in to offer drinks promotion at the same time. It's a completely new scenario combining Big Data with the cloud, mobile payments and the Internet of Things.

"WHAT WE ARE SEEING NOW IS JUST THE TIP OF THE ICEBERG"

WESTERN UNION

Western Union has 70 million customers in 200 countries, enabling 650 million transactions every year, which is around 29 transactions per second. This has contributed to what has been described as "one of the world's largest enterprise data sets".

It has more than 200 terabytes of data, and that data is now growing at a rate of 100 terabytes a year. The data comes in from 174 sources including digital platforms, retail locations, and banking partners.



➤ **CASE STUDY**

JAMIE TURNER

CO-FOUNDER AND CHIEF TECHNOLOGY OFFICER,
POSTCODE ANYWHERE

Jamie Turner is co-founder of Postcode Anywhere – an address capture company that helps organisations obtain better quality data and use it more efficiently. On a website people typically start to input their address with a postcode and then they pick their house number off a list – Postcode Anywhere does that on a very large scale across the world.

We have been going for 14 years and on a busy day we process about 10 million transactions. Our software powers the Royal Mail's postcode finder and the Canadian postal service's postcode finder and we're in talks with several European data owners to do the same thing.

We started collecting search terms and the way customers interacted with the service – more for audit purposes than anything else – but we realised the type of service we could offer customers became more useful as we grew.

Soon we found ourselves with two major pots of data: all the user history going back now almost 15 years – several billion records – and also how our customers interact with us, what pages they visit and what's happened on their account.

That search history lets us know exactly what people get out of the service, and we use that information to improve the accuracy of searches. We take search engine-type technologies and apply them to address and location data sets.

We are comparatively a 'little' Big Data company. We are not like Google or Facebook, which generate terabytes of data every second.

But we have more access to data than the average business and it's a massive opportunity for us. We treat it as a unique asset that's very precious – it's our eyes and ears, and is vital to our future success.

When you rely on automation in this way, it always has to work, and we need 100% uptime. From a simplistic point of view if things don't work, not only do you not get paid, but you also lose credibility and a positive reputation with your customers that has taken many years to build.

Our example is even more unusual because our software is used not only by our customers, but also by our customers' customers, so the knock on impact of downtime, or lack of access, will affect an awful lot of people very quickly.

Our postcode software is used in retail checkouts and the impact in those situations could mean major losses in revenues for our customers, who simply wouldn't be able to process orders for the period of downtime.

To the end user, our business is very simple. You put a postcode in and you get an address out the other end. We have built that business with a reliance on Big Data so quite simply, it has to work all the time, no matter what external issues might put that service at risk.

"WE HAVE MORE ACCESS TO DATA THAN THE AVERAGE BUSINESS AND IT'S A MASSIVE OPPORTUNITY FOR US"

CHAPTER 03

03 LAYING THE FOUNDATIONS

➤ HEALTH SCORE

"Better information means better care". That was the message on millions of leaflets distributed by the UK National Health Service (NHS) in February 2014, offering an opt-out to the new care.data central database of medical records.

Devised as a singular body of information, to give the health service access to information which is otherwise scattered among the UK's 10,981 GP surgeries, care.data was to provide an unparalleled national picture to the medical profession.

"I would say that we are running the health service blind without it," was the argument of one former GP and director of the Health and Social Care Information Centre.

Yet no sooner had the public information campaign landed on the doormats of households across the UK, than it came under attack from privacy campaigners. Criticised as a rushed job which neither the public nor doctors properly understood, it was shelved for a minimum six-month period.

The episode highlighted an inherent conservatism in the public attitude towards the sharing and usage of data. No matter how tangible the benefits — in Scotland, for instance, amputations as a result of diabetes have fallen 40% since a central database of sufferers was created² — there will always be a kickback from those fearful about the use and abuse of personal information.

The fate of care.data also suggested that Britain has some way to go to embrace the full potential of Big Data. And it's not just the scare stories and popular aversion standing in the way of progress. There are serious deficits to be addressed in the skills base and commercial infrastructure before Britain can truly "out-compete, out-smart and out-do the rest of the world" in Big Data, as Chancellor George Osborne called for in his 2014 Budget.

² *The Economist*, 22.02.14, "Caring and Sharing".

A BRITISH MIT

Higher education is where the UK must start, to ensure a pipeline of talent that can fuel a British Big Data explosion.

In many respects UK universities lead the world, evidenced by students who come in their droves from across the world to study at institutions such as Oxford, Cambridge, LSE and Edinburgh.

But even the best universities are well behind the game when it comes to training the next generation of tech talent. There isn't a single UK institution that could hold a candle to MIT, Stanford, Caltech or Berkeley in the States.

The difference is partially one of investment, but above all one of curriculum and connectivity. Take Stanford, the umbilical cord of the Silicon Valley tech ecosystem, providing a ready stream of bright and highly trained graduates ready to staff Google and Apple, or in many cases start their own company.

In the Valley, universities work in perfect harmony with tech companies, and there is an inter-dependence that creates an ongoing virtuous circle between them. It's no different when you look at the fast-emerging clusters in India and Israel, to take just two examples.

What the UK must address is its own fundamental disconnect between universities and business. In both cases, many of the institutions are outstanding, but the dialogue and partnership is not. And that has meant our higher education lags a step behind the rest of the world, teaching the skills that were all the rage a decade ago, but not the technical applications needed by businesses today.

Plumbing in a better interface between business and higher education should be seen as a matter of urgency if Britain is to take advantage of the Big Data opportunity. The experience of a great many UK tech companies is that the graduates on offer are not being suitably prepared by their university courses.

"The graduates that we have here and in the EU, unfortunately, are just not as good as those from universities in countries like the US and India," is the view of Andrew Humphries, co-founder of UK start-up accelerator The Bakery.³

It is fast-growing businesses that do the most to lead innovation in the field, and they do not have the time or resources to train new recruits. The imperative lies with universities to up their game.

Larger employers also have a role to play, co-investing in the training of the next generation of British software engineers. The UK has a proud tradition of developing apprenticeship schemes, but we need big business to get involved.

It is vitally important to involve businesses in the major training decisions, particularly for a sector that will grow as rapidly as Big Data. Employers are the people best placed to judge what training is worth investing in, providing students with real, on-the-job training and fast-tracking them into decent jobs with promising progression prospects.

INFRASTRUCTURE

The Big Data behemoth requires not just a ready supply of people to sustain it, but the infrastructure to cope with levels of information that until recently were considered unmanageable.

Companies increasingly need constant availability of data, and the costs of losing this access – downtime – can be catastrophic. Large organisations running legacy systems can quickly find themselves in the eye of a storm, when software glitches strike.

In December 2013, the government-owned Royal Bank of Scotland saw thousands take to social media after their cards were refused at shopping tills. The previous year, sister company NatWest saw a failed software update lead to chaos, as payments were unable to process from millions of accounts.

Entering into the Big Data game is no small undertaking for companies of all sizes. That means investing in a battle-ready infrastructure for handling unprecedented levels of information.

Without it, disaster can strike, with even a short period of downtime potentially leading to disenchanted customers, complaints and even compensation payouts.

ACTING ON GOOD INTENTIONS

Big Data is well recognised within the UK as a game-changing technology, officially recognised as such by the Department for Business, Innovation and Skills.

All the right noises are being made, from central government to businesses big and small, about how Britain can take advantage of the opportunities it creates.

The Alan Turing Institute, announced in George Osborne's 2014 Budget, is a strong show of intent. But now a news announcement must become a bricks-and-mortar development, a hub that will enable Britain to take a leap forward and make up for lost time in the Big Data market.

With the right talent and software infrastructure in place, there is no reason Britain cannot claim status as one of the world leaders in showing the transformative power of Big Data.

➤ CASE STUDY

UCI

Charles Boicey is the Informatics Solutions Architect for UCI Health, which comprises the clinical, medical education and research enterprises of the University of California at Irvine. *US News and World Report* listed UCI Irvine Medical Center among America's Best Hospitals for the 13th consecutive year in 2013. Earlier this year, the hospital became one of the world's first medical institutions to embrace the opportunity provided by Big Data.

The medical profession has, as with nearly every other service and industry, been transformed by technology. Where we work and the tools we use on a daily basis are a far cry from the hospitals we trained in, let alone the hospitals in which we were born.

The future of medicine will be similarly informed by the revolution in Big Data, with many seeing it as the next frontier in identifying new cures, minimizing the impact of infection, or taking pharmaceuticals to the next level.

Everything from heart monitors and ventilators, from medicine dispensers to thermometers can now send out a seemingly infinite amount of data. And it's been posited that hugely significant trends can be found within all this information, crunching the vast amounts of electronic data that's emitted and collated by the second to extract highly valuable findings.

But a major challenge has been managing the data sets in a way that makes it possible to draw conclusions that are coherent and useful. Some doctors and medical staff still need convincing that Big Data analytics has relevance for the hospitals of the future, let alone hospitals today. Until recently, we at UCI Medical Center would have agreed.

We found that the existing data systems did not allow us to manage the flow of data as it happened in real time, not to mention store the information in its native form.

Most digitally enabled hospitals will be familiar with the current data systems, most notably the Electronic Medical Record (EMR) or Enterprise Data Warehouse (EDW).

They are both useful databases, certainly, but they have their restrictions.

The EMR has to be manually updated, with critical data usually not uploaded until 24 hours after a medical event has taken place. As a result, they can only present findings retrospectively, no doubt useful for when a clinician checks a patient's medical records but not the groundbreaking revolution promised.

Earlier this year, UCI Medical Center became one of the first hospitals in the world to embrace Big Data. We now use Hadoop-based technology to process accurate pattern-set recognitions, use algorithms to monitor patient recovery for non-linear complications, and build predictive-modeling systems to minimize deaths caused by medical error.

This is drastically improving the level of care we offer our patients, with doctors alerted as soon as vital signs cross a key threshold. And it has helped ease the burden on our doctors and nurses, whose heavy patient loads prevent round-the-clock observation.

Hadoop is helping UCI Medical create predictive models that the EMR cannot, especially around heart attack or pneumonia patients. We are now able to detect trends in our patients that we weren't able to before and act accordingly to prevent hospital readmission.

Specifically, we can intercept infections at the earliest possible opportunity, minimizing the risk to the patient. The positive implications this has for patients are vast and I've no doubt more hospitals will follow our lead.

In order to commit to Big Data in this way, we needed to know that nothing will fail – we need round-the-clock availability and reliability of this information every second of every day.

Our healthcare system is on the threshold of a radical overhaul as Big Data starts to make its first tangible mark and shift our approach in a most revolutionary way. Soon we will not be focused on curing people when they are sick, but will be able to identify what can be done to make people well.

³ *Computer Weekly*, 03.09.13, "UK tech startups face hiring roadblock due to immigration policy".

CHAPTER 04

04 THE FUTURE OF DATA

- In many ways the term Big Data will soon be obsolete. As the data explosion grows, today's big will become tomorrow's small.

Research conducted by data storage company EMC found that more data is now transferred across the internet every second than was stored across the entire internet just 20 years ago. The same study reported that the digital universe will be 40 times its current size by the end of the decade, such is the rate of its expansion.

While not everyone is embracing Big Data, there is a strong indication that data-driven approaches correlate with higher levels of productivity and increased profitability.

The *Harvard Business Review* found that the more companies identified themselves as data-driven, the better they performed on objective measures of financial and operational results. Cross sector analysis reported that firms in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and profitable than their competitors.

Elsewhere, a survey conducted by Capgemini and the Economist Intelligence Unit found respondents said the use of Big Data has improved business performance on average by 26%. Almost 60% of the same respondents said they planned to make a bigger investment in Big Data over the next three years.

As more and more business activity is digitised, new sources of information and ever-cheaper equipment combine to bring us into a new era – one in which large amounts of digital information will exist on virtually any topic. The challenge will be to identify the competitive advantage amid all the data.

The financial services sector has begun to recognise the value, with the University of Oxford last year reporting that more than 71% of the industry already uses Big Data analytics. Today's banks are using new data-driven models in an attempt to return to profit following the financial crisis that began in 2008.

Most investment houses used to rely on overnight batch to make trading decisions, meaning their risk management models were forced to use information that soon became out of date. Real time analytics are enabling better trading and risk decisions, safeguarding them against the threat of collapse.

Given the rate at which financial services firms are embracing Big Data, it will be vital to ensure appropriate safeguards are in place to protect against technical failure. When Hurricane Sandy hit the east coast of the US, it ripped right through the heart of New York's financial district. While data was backed up on servers in New Jersey, the neighbouring state was in the path of the same hurricane leaving the financial markets in disarray. As Big Data becomes ubiquitous, events like these will no doubt become increasingly frequent without the appropriate protection.

As with any sector in its infancy, a wave of new companies will no doubt spring up to provide services that deal with problems arising from Big Data – issues and complications that were not anticipated. Just think about the range of different business types to have emerged for the internet.

Not only will the volume of data change, but so too will the variety of sources. Many of the most important sources of data are relatively new, taking the form of messages, updates and images posted to social networks; readings from sensors; GPS signals from mobile phones and more.

Although healthcare and retail have been among the most successful early adopters, the public sector will benefit hugely from better data management.

Tom Heath from the Open Data Institute has said that Big Data can help government cut costs, be more effective and better serve their citizens. During bad weather, a local authority can blend existing data about services, such as road-gritting, with those for at-risk groups, such as 'meals on wheels', ensuring that different providers no longer operate in silo.

Public services are an often overlooked area for innovation. Emergency services have already begun to look to Twitter to aid their efforts, but a more formal social media strategy can provide early warning mechanisms after a major disaster has taken place.

While newspaper reports are divided about Big Data's potential, the wider media has already started to look to audience data to cut through an ever-expanding marketplace.

TV viewing figures have always provided something of a snapshot of how popular a programme is. For years, TV executives placed great weight in the data gleaned from transmitters that monitor which shows are being watched and when. Extrapolated from the data gleaned from a few thousand households, shows are renewed or axed based on the reception given by this small sample.

But instead of a rough estimate of how many people watch a programme, Big Data affords a pinpoint-accurate representation of how long they watched, when they fast-forwarded, what they tweeted. It's claimed that Netflix pushed ahead with its *House of Cards* series because their data found that the BBC original, programmes featuring Kevin Spacey, and films directed by David Fincher to be in high demand among its users.

What's beyond doubt is that companies won't reap the full benefits of a transition to Big Data unless they're able to manage change effectively. Rather than accepting correlations blindly, data scientists will need theory to explain why the patterns are the way they are.

Companies will succeed using Big Data not simply because they have more or better data, but because they have leadership teams that set clear goals, define what success looks like, and ask the right questions.

➤ CASE STUDY

THE ALAN TURING INSTITUTE

In 2014 the British government announced that Big Data was to have a permanent home in the UK. Named after the computer pioneer and Bletchley Park code-breaker, The Alan Turing Institute will focus on new ways of collecting, organising and analysing data.

The government committed to funding some £42 million over the next five years for the project, with universities and other interested parties encouraged to bid to host the institute.

Speaking in his Budget speech, Chancellor George Osborne MP said: "I am determined that our country is going to out-compete, out-smart and out-do the rest of the world."

The announcement was hailed as having the potential to marry the needs of business with the ability of UK academia.

Government expects Big Data technologies will create some 58,000 more jobs in the UK by 2017, contributing some £216 billion to the UK economy.

Science minister David Willetts MP said a focus on Big Data would be crucial to ensuring the UK is able to compete with the world's most technologically advanced countries.

"Making the most of large and complex data is a huge priority for government as it has the potential to transform public and private sector organisations, drive research and development, increase productivity and innovation, and enable market-changing products and services," he said. "The new data research centres will help the UK grasp these opportunities and get ahead in the global race."



➤ **CASE STUDY**

CHANNEL 4

GILL WHITEHEAD
 DIRECTOR OF AUDIENCE,
 TECHNOLOGY AND INSIGHT AT CHANNEL 4

SANJEEVAN BALA
 HEAD OF DATA PLANNING AND ANALYTICS AT CHANNEL 4

Gill Whitehead and Sanjeevan Bala lead the team at Channel 4 that applied Big Data to better understand its audience and reinvent the way in which people consume television.

In January, we announced that 10 million viewers had agreed to share information about how they used the service. This was revolutionary within our industry and something we achieved by creating a value exchange, so when we asked them to give us information we gave them something back.

Across our online offering, we're looking to enhance the user experience, both in terms of the viewer journey and at what points we intervene to make it personal.

Big Data enables a collaborative relationship with clients and end users. We now offer access to our complete archive, exclusive areas and the opportunity to watch premiers – the value exchange must be right.

Amongst all this, we know that to customers Big Data is still new and many are concerned about privacy – this is a challenge to many industries and not just our own. So we created our viewer promise, fronted by Alan Carr, who could explain exactly why we ask for each piece of data and what we use it for.

As part of our viewer promise, we don't want to experiment with anything that could be perceived as being invasive. Our relationship with our viewers remains our number one focus and Big Data must benefit that, rather than risk putting it in jeopardy.

We started our Big Data journey three years ago and at that time our team had two main priorities. The first, and our main focus, was to use Big Data to build and strengthen direct relationships with our audience.

This is something that was never possible before in linear television, but as TV became connected to the internet we suddenly had a vital return path to the consumer.

Our other priority was to deliver a return back to the business via Big Data which we did by creating new advertising products that delivered more targeted advertising opportunities to clients which allowed us to charge a premium and deliver greater returns.

We use Big Data across the business and we now have more direct marketing opportunities that we didn't have before. We're also looking at how Big Data can complement existing research sources to support commissioning and scheduling of TV programmes.



➤ CONCLUSION

"First came the binge, then came the hangover," was the apt summary of a Bloomberg special report on Big Data from June 2014.⁴

The desire to scrutinise Big Data's big promises has led to a fierce debate over the extent of what it can achieve, with critics emerging to claim sighting of a new Icarus flying too close to the sun.

A *Science* magazine report on the much-heralded Google Flu Trends project accused the search giant of "big data hubris", suggesting that the flu-tracking service was delivering consistently wayward estimates of actual infections.

That followed a *Harvard Business Review* article from December 2013, entitled "You may not need Big Data after all", while a *Fortune* magazine column the previous month argued that the tools deployed to analyse the datasets threaten to undermine their credibility.

But as the dust settles, there is little doubt we will see a new wave of companies emerge, designed to help manage the realities of a world built around data-driven solutions – strategies built on intelligent and considered analysis rather than number-crunching alone.

It would be churlish to deny the challenges, but lessons will no doubt be learned from early failures. Despite the recent scepticism, smart businesses will be mindful of the opportunity presented by Big Data. It is providing an unprecedented insight into customers, sectors and trends. And even if it cannot yet track the spread of disease, it is helping hospitals save lives and offer greater levels of care.

Britain must now capitalise on this opportunity. A recent study by EMC found that while the average business will be managing 50 times more information than it does today, the number of IT staff will only rise by 50%. If that holds, then the UK won't be able to fill the 58,000 Big Data jobs that will be created by 2017.

We have to ensure training programmes are in place that will provide the skills needed to cope. This means having universities with the facilities and courses to train the next generation of data scientists and analysts.

The Alan Turing Institute promises to be a huge step in the right direction, but further details remain scarce. We need reassurances that plans are in place to follow up on the Chancellor's promises, particularly with an election looming in 2015.

Creating Big Data apprenticeship schemes will help address this skills shortage in the short term, and we are delighted that WANdisco is the first UK firm to offer these. This is a move we hope will provide the blueprint for other companies to follow.

It's a choice we made because we believe that the Big Data promise must be substantiated by measureable action. Good intentions will not produce the skilled graduate workforce needed to make Britain's Big Data dream a reality, nor will they bring about the required changes in how companies use and think about data. The time for talking has passed, and we must now enter a phase of fiercely urgent action to place Britain at the heart of the worldwide phenomenon that is Big Data.

- British industry will need 1.25 million new graduates in science, technology, engineering and maths by 2020 to maintain current employment markets
ROYAL ACADEMY OF ENGINEERING
- £42 million committed to The Alan Turing Institute
GEORGE OSBORNE MP MARCH 2014
- £73 million additional funding for public sector analytics projects
DAVID WILLETS MP FEBRUARY 2014
- 58,000 more Big Data jobs to be created in the UK by 2017, contributing £216 billion to the UK economy
DAVID WILLETS MP FEBRUARY 2014
- UK Government could save £2 billion in fraud detection, create 2,000 new jobs and generate £3.6 billion in savings
SAS/CENTRE FOR ECONOMIC & BUSINESS RESEARCH 2013

⁴ Bloomberg, "Buried in Big Data".

PREDICTIONS & RECOMMENDATIONS

WANdisco believes that Big Data is good for business. It is transforming the way companies understand their customers, disrupting markets, and has the potential to eliminate discovery by chance.

Britain has an enviable reputation for technological innovation and with the right focus and investment, Big Data will become one of our greatest success stories.

In *Think Big: Britain's Data Opportunity* we have mapped out what needs to be done for the UK to capitalise – here are our key findings.

01

The information explosion fuelling the Big Data movement will only continue to grow as the size of the digital universe expands.

02

Big Data represents this generation's defining moment of competitive advantage – with the potential to disrupt every industry and sector.

03

Both government and business need to get better at explaining the possibilities Big Data presents.

04

The UK faces a technical skills shortage and requires both long-term and short-term solutions to ensure a steady stream of talent to help harness Britain's Big Data opportunity.

05

UK government needs to follow up its commitment to The Alan Turing Institute by making a considerable investment in further education and Big Data apprenticeships.



➤ **ABOUT WANDISCO**

WANdisco harnesses the power of 100% to realise the possibilities of Big Data. Its unique software helps the world's most admired and influential organisations to become stronger, more agile and more competitive, and allows innovators in every field to make the new and important discoveries that will shape the future of society.

WANdisco believes that, in an era of ubiquitous information, an ability to store and query data is the defining factor of success. Our patented replication technology provides organisations with 100% reliable, real-time access to data with no downtime, data loss or latency – making Big Data invincible.

By harnessing the strength, security and stability of 100% certainty, WANdisco unlocks the power of the possible, empowering our customers to push the boundaries of their ambition and benefit from the transformative power of Big Data. Whether to enhance efficiency or protect profits, boost revenue or save lives, with WANdisco one thing is certain: anything is possible.





WWW.WANDISCO.COM