



---

**01** The promise of Big Data

---

**02** Embrace mass customisation

---

**03** Make much better use of unstructured data

---

**04** Exploit the wealth of external data

---

**05** Unleash the power of fast data

---

**06** Turn Big Data into small data

---

**07** Leverage the potential of small data

---

**08** Understand the importance of the final mile

---

**09** Summary

---

**10** Practical first steps

---

---

# 01

## The promise of Big Data

---

**Big Data is one of the most touted and talked about concepts in technology. But is all the hype and hyperbole justified? Does Big Data really deliver value or just leave us all acutely short-changed?**

**To find the answers, let's start at the beginning...**

The term was first coined, about twenty years ago, when large search companies started wrestling with ways to process the huge volumes of data generated by the internet. Driven by the belief – or, to be more accurate, the optimistic hope – that the magic of analytics would someday unlock valuable insights, companies started creating vast data lakes.

And make no mistake, the value of these insights is potentially huge...

An article in Forbes magazine (September 2015) found that a mere 10% increase in data availability, for a Fortune 1000 company, could result in \$65 million worth of additional net income. Furthermore, retailers who leverage the full power of Big Data could increase their operating margin by as much as 60%.

Today, all large businesses and many smaller ones have access to a profusion –and confusion – of historical data. This includes both the information they capture and store themselves and the insights available externally; typically from social media. But are they reaping the rewards promised by all this rich data?

Well sadly, the answer would appear to be a pretty emphatic no! Scratch the surface and the golden prospects offered by Big Data soon turn to the base metal of empty promise. The alchemy just isn't working. Despite the fact that we have created more data in the last two years than in the entire history of the human race, less than 0.5% of this information is ever analysed or exploited.

But why? What can be done? And how exactly can companies turn lost opportunity into tangible business value?

**We believe the answer lies in seven key steps:**

- a. Embrace mass customisation
- b. Make much better use of unstructured data
- c. Exploit the wealth of external data
- d. Unleash the power of fast data
- e. Turn Big Data into small data
- f. Leverage the potential of small data
- g. Understand the importance of the final mile.

---

# 02

## Embrace mass customisation

---

**It is surprising, given how much data large companies have on us, that they seem to have so very little awareness of us as individuals. They seem only to view us as part of a generic market segment or to play back to us our own behaviours in overly-simplistic ways. In today's customer-driven world this simply isn't good enough.**

In their definitive book on mass customisation - 'The New Age of Innovation' – the co-authors (C.K. Prahalad and M.S. Krishnan) say:

*"There is a fundamental transformation of business underway. Forged by digitisation, ubiquitous connectivity and globalisation, this transformation will radically alter the very nature of the firm and how it creates value... value is based on unique, personalised experiences of consumers. Firms have to learn to focus on one consumer and her experience at a time, even if they serve 100m consumers. The focus is on the centrality of the individual".*

This is the central challenge of Big Data: the future belongs to those companies that use information to personalise the experience of each and every customer.

---

---

As more and more services become digitised and commoditised, customer loyalty will be increasingly driven by the data they have shared with a company. This bond strengthens with every customer interaction as does the quality of data. The really clever companies are the ones that use this rich intelligence to simplify and focus their services – creating a uniquely individual customer experience.

There is no simple path to customisation, but the essence is to organise data and insights around individual customers, moving them from 'unknown', to 'partially known' and, ultimately, to 'known' as they move through the customer lifecycle:



Unknown



Partially Known



Known

---

### 03

## Make much better use of unstructured data

---

**Many companies appear to be more preoccupied with capturing and storing data than harvesting, analysing and exploiting it. Inevitably this has created vast lakes of untapped information where concerns over quality, inconsistency, compatibility, security and privacy have constrained usage. Unnecessarily!**

This problem is often compounded by an over-emphasis on storing structured data and a failure to recognise the power of unstructured data...

*One of our Clustre member firms has a powerful tool which analyses and organises unstructured data using Latent Semantic Indexing. Developed for the state security sector by one of our most secretive 'skunk works' teams, this tool can perform conceptual search and multi-lingual analytics without the need for Natural Language Processing.*

And you certainly don't have to limit yourself to unstructured text...

Video, for example, is more difficult to process but it is often far richer in information content. Today's really savvy companies are making increasing use of video as a means of communication – both internally and externally – with suppliers and customers. Indeed, the real innovators are using video as a medium for capturing, analysing and disseminating data.

*Another of our new breed of innovators is exploiting the power of smartphones to capture hugely revealing, real time customer insights. It uses the video selfie to engage with a pre-*



*selected group of consumers - capturing their experiences at the very point of sale...*

*Sentiments and semantic themes can be analysed, body language can be studied and advanced algorithms can identify key learning points.*

*Most revealing of all, this technology is closing the chasm between customers and senior management. It is bringing the real 'Voice of the Customer' into the Boardroom.*

And another of our member firms has developed an app that allows field and sales teams to record meeting notes via structured video for sharing and knowledge transfer. And, to prove just how powerful this can be, the next case study shows the impact this app had on a Fortune 500 client...

---

## Case study: Capturing Global Information

With \$12.4 billion of annual revenues, this client ranks among the world's fastest growing and highest performing companies. Quite simply, it is one of the most remarkable global success stories.

But with success has come a whole raft of new challenges...

Following a period of rapid growth and acquisition, the inevitable stretch marks began to show. Less than 50% of sales opportunities were being recorded in a timely and accurate process. This was having a damaging impact on forecasting and resource planning across the entire global group.

To find the solution to this global information-capturing problem, management turned to one of our innovation firms. Aware that sales professionals are notoriously reluctant to fill-in forms, our firm helped the client to leverage a combination of radical technologies that transformed the data harvesting and profit performance of this Fortune 500 company...

Working in very close collaboration they:

- Turned reluctant sales teams into avid adopters of radical technology
- Put mobile technology at the heart of this global transformation
- Used innovative guided video capture technologies to accelerate adoption
- Achieved sales strategy alignment... greatly enhanced information flows... and significantly improved win rates and forecasting accuracy.

## 04

### Exploit the wealth of external data

**Companies often confine themselves to their own customer and transactional data without realising how much more data is available from external sources.**

Quite apart from the mass of data generated by a company itself (through its direct interactions with customers), there is a virtually unlimited amount of additional data available via the internet. Social media streams are one of the richest sources of consumer insight and the analysis of these streams – social intelligence – forms a critical and fast-growing part of the big data puzzle.



Social intelligence can be used to surface the unmet needs and frustrations of customers. These are the true signposts to innovation opportunities and profits:

*One of our member firms – a world-leading innovation consultancy - has developed a unique system for analysing social media streams and identifying hotspots of consumer frustration. These insights are then used to create product innovations that address these specific consumer needs. The system can even model the precise impact of these product innovations – accurately calculating the increased level of consumer satisfaction and the growth in profit performance.*

And the wider sources of information available on the internet can be used to augment other customer intelligence - providing a more complete, rounded and timely level of insight:

*Another of our Clustre member firms provides B2B sales intelligence that mines data from a myriad different sources. Leading financial service companies were among the first to appreciate and adopt this service. Better informed sales professionals now have intuitive, real-time insights into their customers' business. What's more, this service integrates seamlessly with existing CRM systems - supercharging account development and delivering dramatic improvements in sales force effectiveness.*

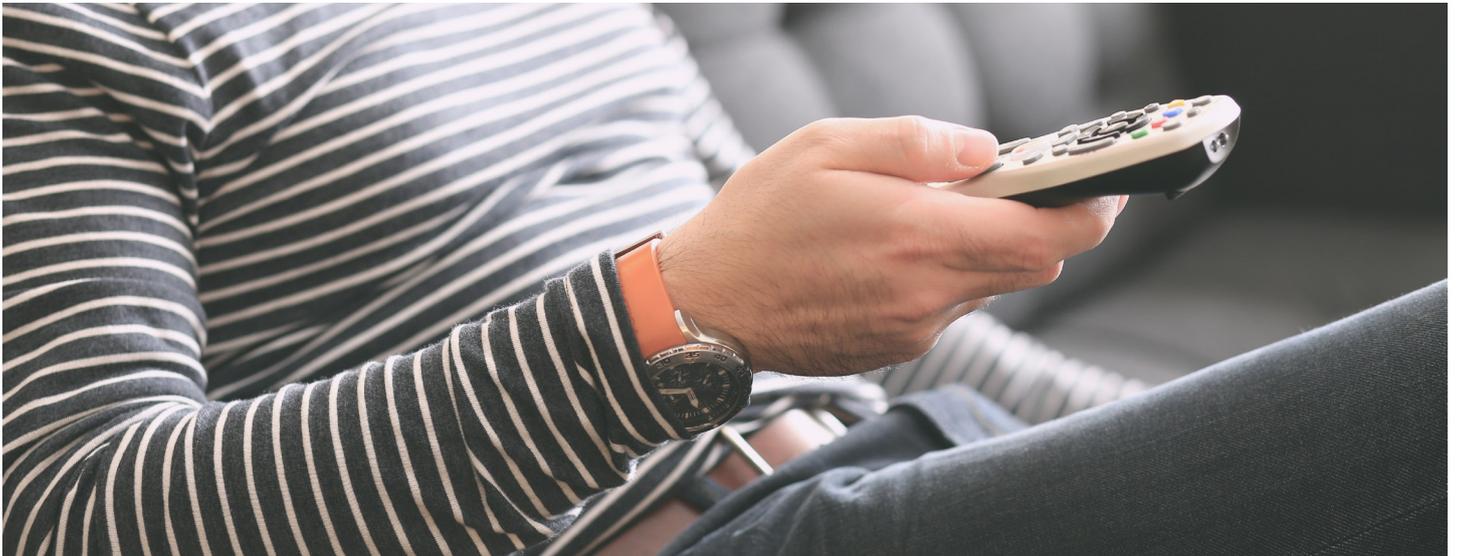
**We live in a world of constant and continuous connectivity. A world where smartphones have delivered the promise of instant communication...**

**But that promise comes at a price.**

The 24/7 pressure to react to data feeds is insistent and incessant. To compete effectively and to influence critical buying decisions, we simply have to respond in real time. And it's very clear that slow-processed historical data simply cannot compete in this arena. This is where the rapidly emerging technology of fast data is now flexing its muscles.

Apache Spark uses in-memory processing to deliver advanced data analytics in orders of magnitude faster than Hadoop. But, to fully exploit the power of fast data, you have to architect systems in a new way. Often referred to as 'reactive' architectures, these new system-building methods are 'turbo-driven' by micro-services. Highly responsive, resilient, elastic and message-driven – this is the future for fast data.

One of our member firms (the creators of the Scala programming language used in Apache Spark) has extensive experience with reactive systems and fast data technology. This team specialises in implementing reactive, real-time solutions to Big Data problems – and, to illustrate the value it delivers, here is a typical case study:



**Case study:  
Video on-demand**

*This Clustre firm works closely with an American cable company that has recently introduced an on-demand video service. The introduction of pay-on-demand viewing certainly opened up massive new opportunities for profit; but it also created equally massive data headaches...*

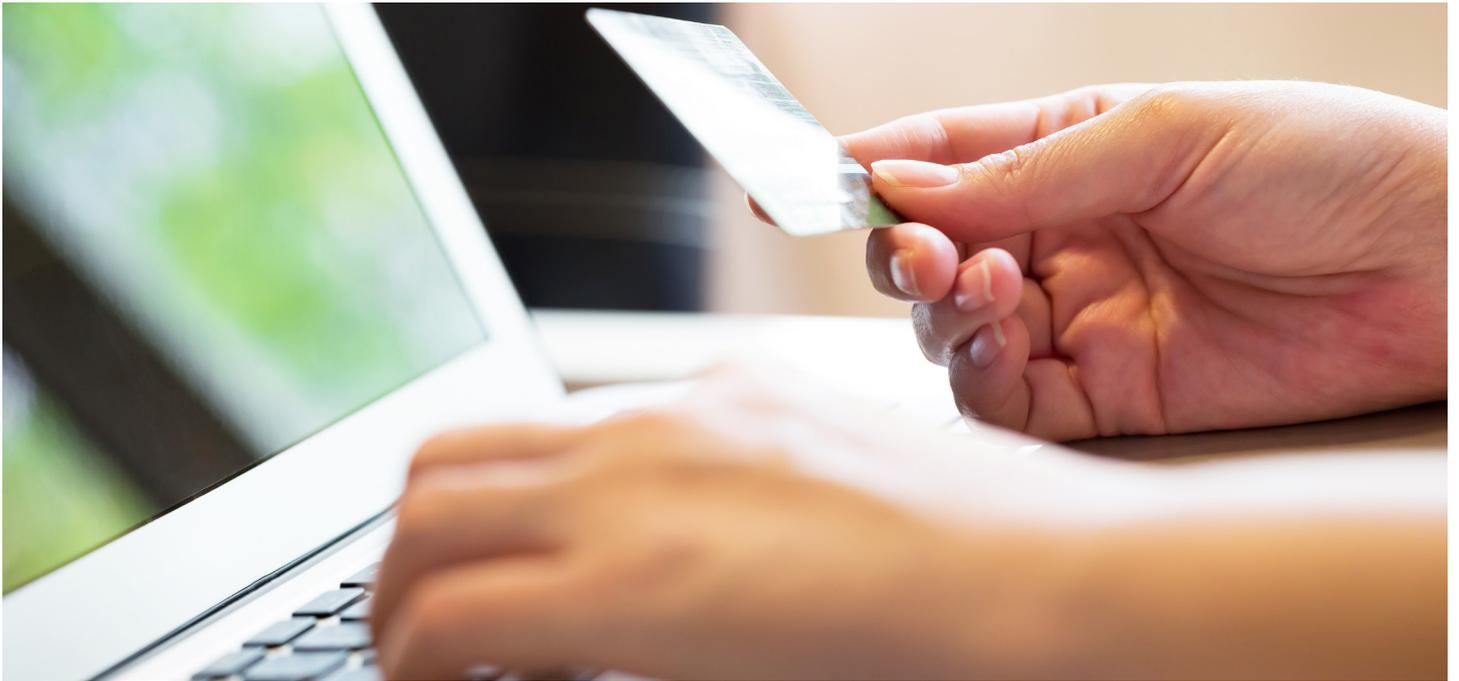
*The cable company operated a large relational data store containing all customer information. It had worked well until the arrival of on-demand – but then everything instantly changed. The cost of having one data store to determine what customers could watch at any given moment suddenly spiralled. The company had to find a way of splitting the functionality – so they turned to our innovation specialists for an inspired solution...*

*The priority was to turn the relational store into a write-only view and to create a separate readable view that would cover all the watchable entitlements for different devices. Admittedly, this did involve some essential data duplication but this was quickly accommodated.*

*Synchronisation is critical to large cable companies. It is essential that customers can watch the programmes they have paid to see... but it is equally important that non-subscribers are denied access to programmes they are not entitled to view. To solve this complex customer challenge, the 'read' store was configured so that it could constantly interrogate the 'write' store. It instantly recalculates so that any omission on the 'read' view quickly self-heals to provide a seamless service. Now that is very clever.*

**We have become so preoccupied with capturing huge amounts of historical data that we have lost sight of one very key fact: humans are not programmed to process vast volumes of information. We are simply not wired to be Cray computers.**

**Consequently, we need to slice-and-dice Big Data to create digestible portions of small data. Chunks of focused information that will directly impact and improve business performance. Let us put this in context with a global example:**



#### **Case study: eBay Europe**

*eBay Europe commissioned one of our Clustre member firms to tackle the challenge of listing optimisation for their growing community of professional sellers. Essentially, eBay needed to cut through the noise that sellers face every day and provide a definitive tool to optimise sales performance.*

*Admittedly, there is no shortage of tools available to 'help' eBay sellers. The online shelves are creaking with self-help manuals. The snag is, most are full of worthy but largely theoretical graphs and charts. Few (if any) provide definitive evidence-based recommendation – practical action-points drawn from precise seller experience.*

*The Clustre member firm's solution was to create a customer recommender scheme specifically for eBay sellers. A sorted hierarchy of actions that are ranked by effectiveness and constantly updated to reflect the subtle changes in customer buying habits/behaviours.*

*And to put this solution in context, here are just a few of the engineering challenges that were resolved:*

- *Daily data collection for more than 50 million listings in the UK and Germany alone*
- *The development of a custom Product Taxonomy for listing comparisons across multiple sites and data structures*
- *The merging of all listing and performance data supplied by eBay*
- *The painstaking analysis and comparison of top-selling listings in each category*
- *The identification of all data gaps, missing keywords and pricing insights.*

**Big Data is a massive misnomer. Many people believe that meaningful insights can only come from vast quantities of connected data. Indeed, many companies cling to the belief that, to be better informed, they have to collect ever more data.**

**Nothing could be further from the truth because, in reality, less is invariably more. Very often the most valuable intelligence is drawn from the sparsest of data...**

**To prove that point, check out this fascinating case study:**



### **Case study: Angel Trains**

*Another of our Clustre member firms was commissioned to keep track of critical rolling assets...*

*Railway engines are extremely costly and complex assets. And many of those complications stem from the number of different stakeholders involved in running our railways...*

- *For a start, there are the train owners – typically finance companies – who bank-roll each locomotive's purchase...*
- *They lease the engines to operators who hold the franchise to operate particular UK routes.*
- *For convenience, many operators sub-contract the servicing of their leased engines to third party contractors. So, effectively, neither the train's owner nor the operator have any direct control over the care and maintenance of this costly asset.*

- *And to complicate matters further, the trains run on tracks owned by Network Rail. So the owners do not know how many miles are covered by any specific locomotive over any given time.*

*This paucity of data raises very real concerns. The owners know next to nothing about the state of their assets. They have no way of knowing if operators are taking good care of their costly trains or if they are keeping to their contractual terms and conditions. So they have no idea whether their depreciation assumptions are wildly optimistic, overly pessimistic or bang on the money.*

*Ten – or even five years ago – this dearth of data would have derailed attempts to find answers. But modern machine-learning techniques can now provide outstanding insights with relatively small datasets... and we mean very small!*

The information on each train engine was virtually microscopic. The only real dataset available to owners came from the locomotive's on-board, fuel-tank sensors and GPS. Every five minutes, these transmit information on changing fuel levels to a central data store. So, in theory, it was possible to track the levels of diesel consumption for each engine over a measured timeline.

In addition, Network Rail's own app reveals when trains depart and arrive in stations.

Our innovation specialists tapped into these twin sources to establish train movements. But there was a snag... Angel Trains do not know which particular locomotive is operating at any particular time on any specific scheduled service.

But in the absence of information, machine-learning – the science of inference – suddenly came into its own...

The first task was to clean the data because the fuel tank readings were pretty 'noisy'. The read-out graph of the diminishing diesel levels showed big spikes and troughs caused by fuel surges as the train corners, accelerates and brakes.

To remove this noisy time series, the team used an L1 Kalman filter. Then, using a support vector machine, they started to look for recognisable patterns. Focusing on small datasets, they were able to identify time zones when, for example, the fuel level rose vertically it was almost certainly on a stationary refueling stop. Other periods show the train to be in 'Hotel Mode' – stationary but still using fuel to heat and light carriages.

Forensic deductions. The analysis then became even more forensic. The team studied the train's movements over a particular time. From this data it was possible to deduce stationary periods for the train to be recommissioned or refueled... running periods between stops... the frequency of stations... the distance between these stations... and even the probable route and type of service (local, commuter, express etc.).

Engine usage. With the use of standard machine-learning techniques and the support vector machine, a very credible and revealing picture of train usage was emerging. Trains with less than 10% movement were deemed to be 'unused'. Engines with under 50% movement were considered to be 'lightly used'. Locomotives with more than 50% movement were classified as 'heavily used.'

Fuel efficiency. They then turned the spotlight on the critical issue of fuel efficiency. By classifying the data in a different way, the stats showed exactly when and where there was high fuel usage (more than half a litre per minute), medium fuel usage and low fuel usage.

Engine matching & intelligence. The team then used weighted maximum matching algorithms to discover which trains had been paired for different routes (typically trains are driven by two engine units, front and rear, with carriages in-between). By aggregating the data, the team discovered that certain engines are routinely paired. Another valuable nugget of operational data.

Mapping journeys. Finally, it was possible to map engine journeys with a high degree of accuracy. By matching the pairings classification to Network Rail's movement data – once again using the weighted matching algorithm – the team tracked one train throughout a hard-working day. Leaving Harrogate at 07:34 it travelled to Kings Cross, then on to Newark North Gate, before returning to Kings Cross and finishing the day in Sunderland.

That was quite a journey – and it all started with a simple fuel reading!

### **The take-home message is crystal clear:**

You do not need large chunks of data to undertake meaningful automated analytics. These fuel time series are literally only a few kilobytes per engine but they generated insights that would be impossible to re-establish manually. Even on very small datasets, machine-learning techniques outstrip and outperform anything that humans can deliver.

## 08

# Understand the importance of the final mile

**In almost all cases, it is the “final mile” of the Big Data journey that counts. The point when data is distilled into something that can be easily understood and acted upon (without drowning the recipient in data!).**

The interrelationships, correlations and mappings of large datasets and their slicing-and-dicing into meaningful intelligence can often only be understood through visual representations...

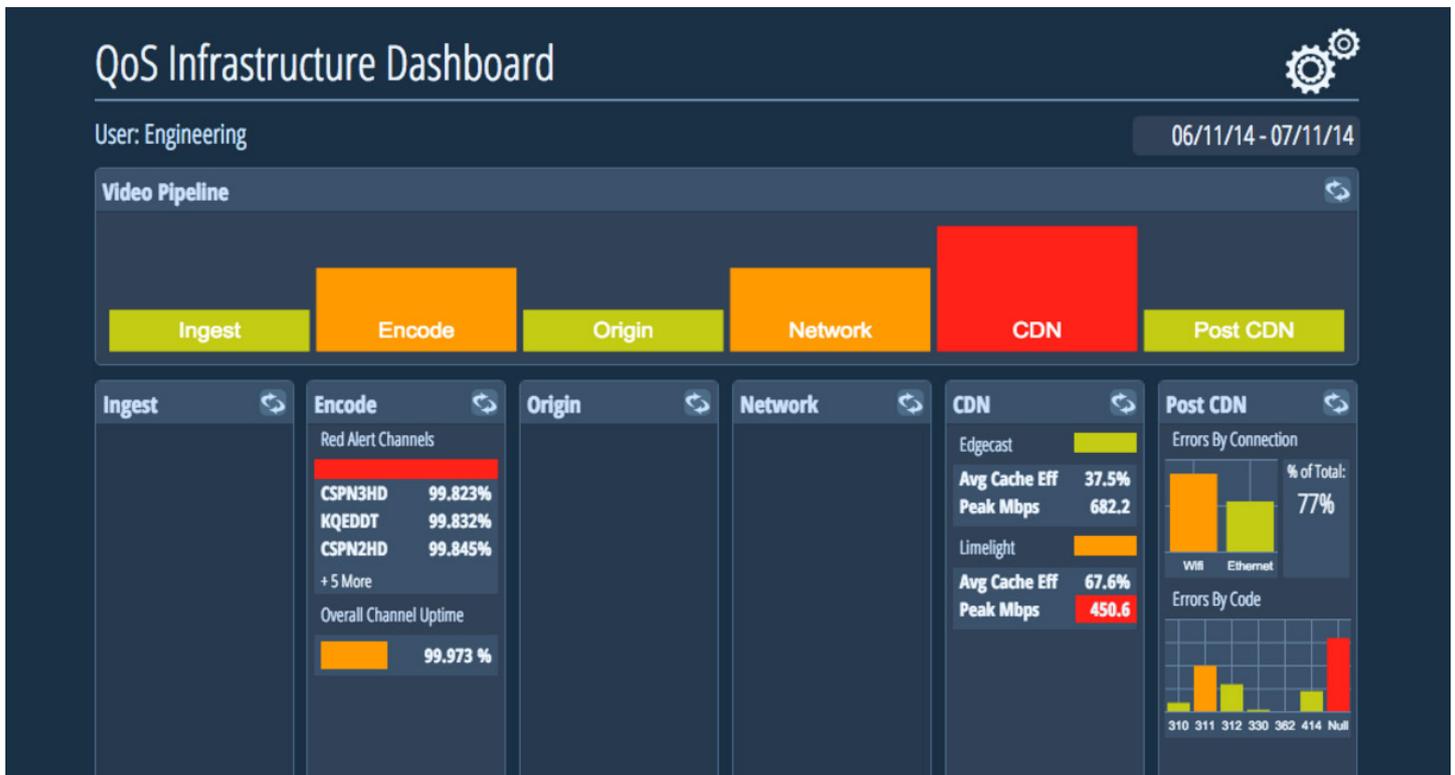
Now there are many standard tools available for generating these visualisations and, sometimes, these are perfectly adequate. However, many organisations fail to grasp the power of graphic simplicity. Instead of visual clarity, their reports and presentations are confused and overly complicated. It's a lost opportunity to cut through the clutter and bring key messages to life with simple imagery.

Once again, less delivers more. For genuinely customised insights it is often necessary to create bespoke visualisations. Here is one example of a complex analytical project where visualisation helped to illustrate the key learning points...

*The client's very specific goal was to assess the performance of one particular product on a national level...*

*The product was comprised of a series of elements which had a direct impact on the overall quality of service. This was not a trivial exercise.*

*The team's approach was to break everything down into a pipeline of solutions – each one solving its own set of questions with a custom visualisation. This gave the client a high-level 'State of the World' perspective on each step in the pipeline – but it also offered the opportunity to drill down into any specific service issue... for example, in the Content Data Network (CDN) used or their own network performance.*



---

## 09

### Summary

---

**The majority of large organisations are doing something with Big Data. Most are diligently constructing 'data lakes' and building pipelines to fill them. But it is questionable whether many of them know how to leverage and exploit this growing asset.**

On one level, it probably doesn't matter; data lakes are relatively cheap to construct so it seems sensible to store information rather than profligately throw it away. However, on another level, we do need to ask some serious questions about the value Big Data can deliver.

Big data is, perhaps, the ultimate example of a solution looking for a problem. Having put in place processes to capture and store these mind-numbing volumes of data, companies now need to work out what to do with them.

This shouldn't really come as a surprise and is actually an opportunity for some genuinely creative thinking. As with so much in business, the trick is to start with the real problems and ask how all this data can be used to do business in a way that is different from – and better than – the past.

**So where do you start?**

---

## 10

### Practical first steps

---

**Try these simple, very practical, first steps to address this challenge and exploit the opportunity of Big Data:**

- a.** Identify opportunities to use Big Data to improve your customers' experience.

Start by identifying real customer problems and points of frustration in the customer experience. Look for things that will make a real difference rather than just 'nice-to-haves' – things with a real wow-factor that turn a customer's frustration into genuine joy.

Wherever possible organise data and insights around individual customers and embrace the concept of mass customisation.

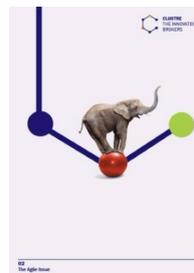
- b.** Focus on 'fast real time data', not just 'big historical data' – develop real time insight from real time data streams.

In order to make a powerful difference, the impact of Big Data/fast data analytics has to be felt by the customer at the point and moment of interaction. Increasingly, this means mobile - so it is vital that you understand and leverage the value of real time data streams from social media.

- c.** Refine Big Data into small data. Make Big Data smaller, more focused and much more actionable. And don't be deterred if you don't think you have enough data to start with!

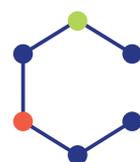
The key to driving business value from Big Data is to distil actionable information and insights. Only then can they be applied to individual interactions with individual customers. And don't forget the importance of that final mile... the power of simple visualisation.

- d.** Use agile development techniques to rapidly prototype and develop customer-relevant solutions.



**This is a great place for trial and error experimentation. To inspire you to develop and evolve your own solutions, check out our Agile PoV: [www.clustre.net/pov-agile-2/](http://www.clustre.net/pov-agile-2/)**

---



**CLUSTRE**  
THE INNOVATION  
BROKERS