

# RESEARCH PAPER

## Big data – are we nearly there yet?

A look at the degree to which big data solutions have become a reality and the barriers to wider adoption

May 2013

Sponsored by



## **CONTENTS**

Executive summary	<b>p3</b>
Introduction	<b>p3</b>
Attitudes to data	<b>p3</b>
Who is driving big data?	<b>p5</b>
Changing data types	<b>p5</b>
Quality versus quantity	<b>p7</b>
Barriers to big data	<b>p8</b>
Conclusion	<b>p9</b>
About the sponsor, Intel	<b>p10</b>

This document is property of Incisive Media. Reproduction and distribution of this publication in any form without prior written permission is forbidden.

## Executive summary

Like the data itself, the big data bandwagon seems to be travelling at such a pace that clambering onboard is becoming an intimidating prospect. Many business organisations seem unsure about the benefits that big data can bring and are engaged in overcoming more immediate data challenges.

*Computing* undertook comprehensive research consisting of an extensive online survey, a series of face-to-face interviews and a focus group forum, to understand exactly what business organisations really think about big data and what it means for them. The research explored how attitudes to data are changing and what challenges businesses are facing in their attempts to gain ever more insights from the increasing quantities of data generated, both inside and outside of their organisations. We asked who was driving big data projects and why objections to big data may be occurring. The paper features detailed discussion of the findings and the apparent gap between the immediate data challenges facing business organisations and the aspirations of big data vendors.

## Introduction

Discussion of big data and what it represents is taking up ever-increasing amounts of space in the IT press. It is a hot topic for analysts too and predictions for the growth of the big data market are indeed impressive. IDC, for example, predicts a compound annual growth rate of 31.7 percent with revenues hitting the \$23.8bn mark by 2016.<sup>1</sup>

However, as is often the case, the vision of big data held by vendors, analysts, bloggers and other thought leaders is very different from the reality in many business organisations. Definitions of big data vary but for the purposes of this paper it will be defined as a collection of technologies to enable the capture and real-time analysis of large volumes of structured and unstructured data for the purposes of business intelligence. While it is undisputable that the volume of data being generated is growing exponentially and that an increasing proportion of this data is unstructured and generated externally to organisations, businesses are taking their time to decide how best to make use of this data.

## Attitudes to data

*Computing* asked organisations to choose some words to describe what their organisation was like today in terms of their attitude to data management. We then asked them to choose from the same list of words to describe their organisation's stance five years ago. It is clear from the results that attitudes to data have changed significantly in that time. The most frequently occurring words used to describe the picture five years ago were "Fragmented" (35%), "Analysis Poor" (31%), "Inconsistent" (29%), "Slow" (29%) and "Disorganised" (27%). Only 13 percent rated their organisation's attitude as competent and 12 percent as customer-driven.

However, the most frequently occurring descriptions given to describe current attitudes were "Technology driven" (34%), "Governance focussed" (31%), "Process driven" (30%) and "Customer driven" (29%). This finding illustrates changing attitudes towards data very clearly, and the answers given by our respondents when we asked them to choose words to describe how they envisaged their organisations to be managing data in five years time drive the point home.

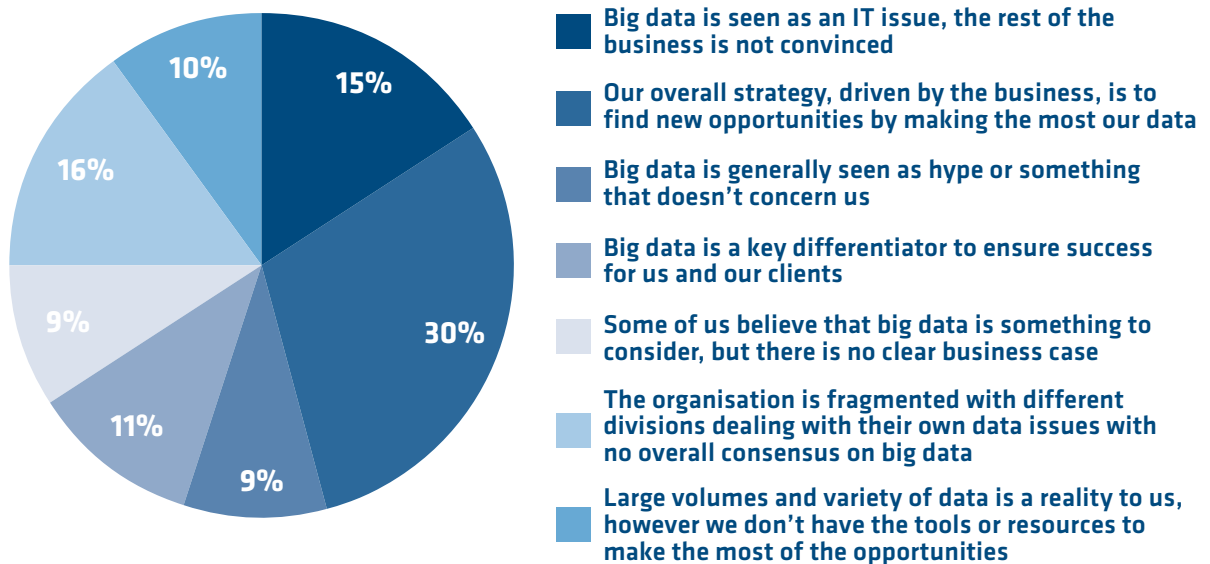
<sup>1</sup> <http://www.idc.com/getdoc.jsp?containerId=prUS23900013#.UWkMv0rfREM>

## Big data – are we nearly there yet?

The largest proportion of respondents (42%) stated that “Analysis rich” best matched their picture of data management in the future, followed by “Efficient” (41%), “Organised” (39%), “Listening to Business Needs” (37%) and “Controlled” (36%).

At first glance the answers to these questions provide a very encouraging picture of attitudes to data management falling into line with that envisaged by big data vendors. However, the picture is not quite as rosy as it first appears. *Computing* asked: “Which of the following definitions is the closest to describe your organisation in terms of dealing with (big) data at the moment?” The results are shown in Figure 1.

**Fig. 1 : Which of the following definitions is the closest to describe your organisation in terms of dealing with (big) data at the moment?**



The largest proportion of those responding to the question (30%) gave the response: “Our overall strategy, driven by the business, is to find new opportunities by making the most our data.” This is encouraging, but the problem of fragmentation of responsibilities becomes apparent with the second and third greatest proportion of respondents stating that “The organisation is fragmented with different divisions dealing with their own data issues with no overall consensus on big data” (16%) and “big data is seen as an IT issue, the rest of the business is not convinced” (15%). Only 11 percent saw big data as a key differentiator to ensure success for themselves and their clients.

The problem of big data being seen as an IT issue is a very real one and merits further analysis.

## Who is driving big data?

The fact that such a significant proportion of respondents chose the phrase “technology driven” to describe their current attitude indicates that any big data projects are being driven by IT teams rather than the areas of the business that can potentially derive the most benefit such as finance and marketing. Technology needs to enable the business but until the business is driving these projects they will struggle to get off the ground.

 **Until the business is driving big data projects they will struggle to get off the ground**

In order to provide a more detailed analysis of how big data projects are coming to life, *Computing* asked: “Where did the initial demand to investigate big data solutions come from?” By far the largest proportion of respondents (43%) stated that the initial demand for big data had come from the IT department. Significantly smaller numbers stated that the impetus for their big data projects had come from marketing (8%), sales (6%) or an evangelist in the boardroom (9%).

While it is to be expected that those who work most closely with technology are going to be excited by the transformative potential of big data technologies, the fact that the departments who potentially have the most to gain from this technology are not driving its adoption should be a cause for concern. Some insights into why this is the case came from our focus group participants.

The following quotation comes from a participant based in education: “I think sometimes the business doesn’t know how to ask. They don’t know what they’re missing so they don’t ask the question until we give them something and then they say ‘oh hey, this is very helpful’. So in a lot of ways, IT drives the future.”

Another participant based in the service sector indicated that the problem may be partly one of perception: “Giving it the label ‘big data’ makes it sound like an IT issue and that doesn’t help...”

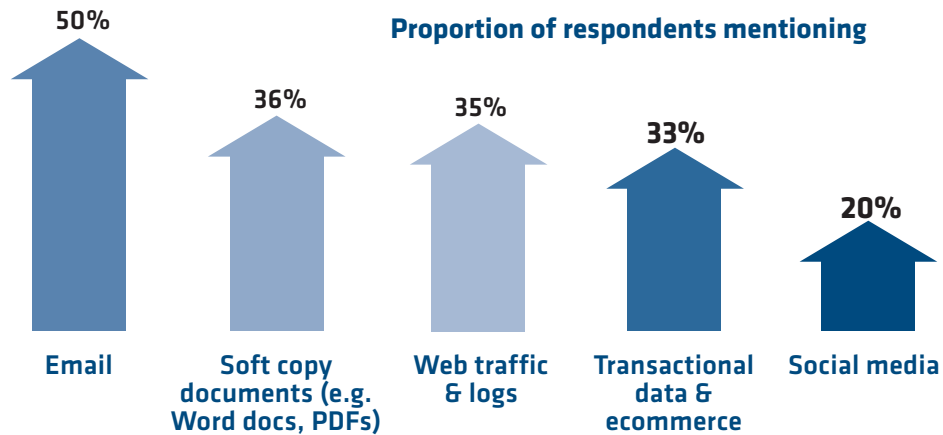
## Changing data types

As we discussed earlier, the growth in data volumes is seemingly unstoppable. However, discussing the growth in volumes is not in itself particularly meaningful. Data volumes are always increasing. The growth in specific types of data is more interesting.

*Computing* asked survey respondents to select their three most rapidly growing data sources. As can be seen in Figure 2, the fastest growing area cited was email (50%), closely followed by soft copy documents (36%), web traffic and logs (35%) and transactional/ecommerce data (33%). Social media, and the need to draw useful insights from it, is an integral part of the big data discussion, but the number of respondents citing social media as one of their three fastest growing sources of data was fewer than 20%.

## Big data – are we nearly there yet?

**Fig. 2 : Main sources of data showing marked increases in volume**



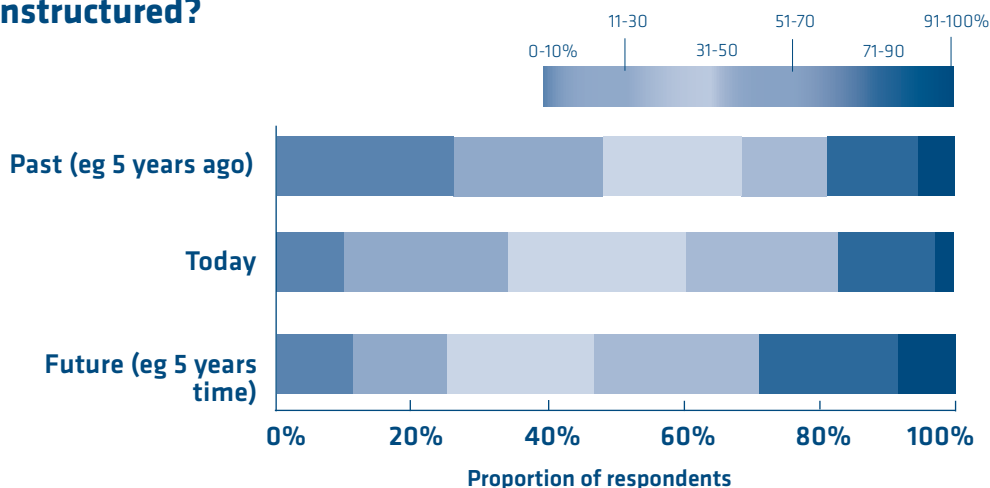
*Computing* also asked about domains of data. Approximately 60 percent of those responding to our survey stated that customer data was their fastest-growing domain. There is a large gap between that and the next category – financial information. Only 34 percent cited this as one of their top three fastest-growing domains and 23 percent mentioned product data.

### Customer data is growing faster than any other domain

It is customer data that is growing faster than any other domain and it is this customer data which business organisations have to mine more efficiently to provide the intelligence necessary to retain and consolidate their market positions.

The challenge facing businesses in making sense of this huge amount of customer data is made greater by the fact that so much of this new data is unstructured. Examples of unstructured data include voice calls, video, handwritten records – basically anything that does not fit into a relational table. *Computing* asked what percentage of data collected by our respondents' organisations was unstructured, and how this proportion compared with the picture five years ago. *Computing* also asked respondents to indicate how they believed this proportion would change in the next five years (Fig. 3).

**Fig. 3 : What percentage of data collected by your organisation is unstructured?**



The results indicate that the shift towards unstructured data appears to be gathering momentum. Almost one third of respondents believe that unstructured data will constitute more than 70 percent of all data processed in the next few years, whereas 15 percent say that is already the situation today. Furthermore, 40 percent of respondents to our survey stated that between 31 percent and 100 percent of their data is externally generated and a majority of respondents expected this proportion to rise in the next five years.

The integration of structured and unstructured data is a key challenge for many of our respondents.

## Quality versus quantity

The increasing quantity of unstructured data being generated and stored is simply part of the trend to store more and more of all types of data. This growth in storage requirements is part of what is driving the growth in big data as a whole. As storage has increased in capacity and reduced in price, the trend has been to store ever-increasing amounts of data. This is happening for many reasons. In the case of the financial services industry, regulations have become increasingly stringent since the sector became subject to intense scrutiny following the crash in 2008. All forms of data such as voice and video fall within the remit of the regulations.

However, even in industries not subject to such tight legal and industry regulations, there is still a growing tendency to hoard data with the hope of doing something meaningful with it at an undefined point in the future. A “just in case” mentality is firmly entrenched in many organisations when it comes to data retention. One of our focus group participants based in manufacturing stated: “If they don’t know how to use it now, they will in the future... The more you hoard, the more of a timeline you achieve, then the more detailed your profiling becomes...”

However, just because you can store ever-increasing quantities of data does not mean you should. More than one of our respondents raised concerns about data provenance. Data volumes may be escalating but the quality is not. Accumulating ever-increasing quantities of poor-quality data is, at best, pointless. Finding the good data and eliminating the bad remains the first part of the data challenge for many organisations.

What was also striking among our survey respondents and focus group participants was the fact that even though they were storing more and more data, they were not gaining any meaningful insights from it.

As one participant put it: “We’ve got lost in the morass of data and where we are needing to mature is in the meta data. It’s because we get lost in the morass of data that we end up spending an inordinate amount of money and time, manually stitching together disparate databases... If you have meta data then that stitching together can happen in a more automated way, hopefully, and start to alleviate some of the manual effort which is just based on joining things together.”

Big data thought leaders get understandably excited about the transformative potential of real-time analysis on ever-larger quantities of data. However, a collective step back allows us to see that the immediate data challenge facing businesses is a tripartite one of maintaining quality, integrating different types of data and being able to analyse that data to a sufficiently rigorous extent.



**The immediate data challenge is a tripartite one of maintaining quality, integrating different types of data and being able to analyse that data to a sufficiently rigorous extent**

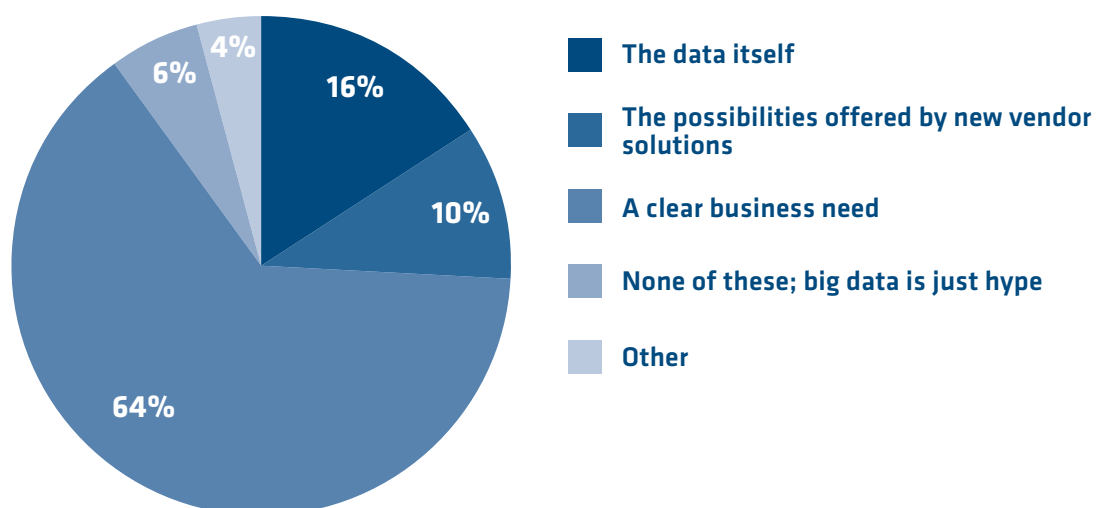
# Barriers to big data

*Computing* asked: “What are the main barriers to implementing big data solutions in your organisation? The single largest proportion of respondents (42%) stated that their biggest problem was the cost. Thirty-eight percent stated that a business case for big data was absent. These two factors are of course related. The perception of big data as a large, costly project doomed to run and run without delivering the promised benefits seems to be hovering over many organisations.

However, the following two largest barriers to big data are more technical. Disparate systems and data locations were cited by over one third of respondents (36%) and a lack of skills was also a problem for approximately one quarter (26%). Only slightly fewer were troubled by security concerns and the problem, already discussed, of linking structured and unstructured data together.

The picture drawn by our research is one of a stand-off between big data vendors offering real-time analysis on huge streams of different data types and business organisations wanting solutions to help them make more meaningful use of the data they already have. When asked “What would be more likely to make your company more interested in implementing a more advanced (big) data solution?”, a massive 64 percent of those responding chose: “A clear business need” (Fig. 4). It is clear that, at present, organisations just do not see a business case for many of the big data solutions available and they view it as an all or nothing project.

**Fig. 4 : What would be more likely to make your company more interested in implementing a more advanced (big) data solution?**



The way forward is to start small. Business organisations can be making a start on big data before total buy-in is achieved, starting small and keeping the goals clearly in sight.

Community versions of Hadoop distributions are often free to download and some now offer enterprise-grade security and performance together with management tools for Hadoop clusters and many other features to ease deployment of Hadoop in the enterprise. Such distributions are ideal for getting started and provide an upgrade path once the decision has been made, as part of a phased approach.



Taking a step-by-step approach is the best way to discover what additional resources will be needed as well as playing an important role in convincing business decision makers of big data's validity.

The last word on how to make big data a reality should go to our focus group panellists, some of whom have implemented or are in the process of implementing their own big data projects.

“Start small, have a base and then work up with them. Bring in help from outside if you need to.” (Services panellist)

“My advice to anyone thinking of doing a big data exercise would be to go for it because the benefits are quite enormous.” (Local government panellist)

“Ensure you have got key stakeholder alignment around what the outcome is going to be because the allure of data to an untrained person means that you open a whole rabbit warren and the insights that you want just don't exist.” (Media panellist)

## Conclusion

Attitudes to data in business organisations are evolving fast from “fragmented”, “analysis poor” and “inconsistent”, to “technology driven”, “governance focussed”, “process driven” and “customer driven”. “Analysis rich” was the most popular aspiration for the future. However, despite the multitude of benefits which big data can bring to marketing, finance and business operations, big data is still perceived as an IT issue, with IT teams driving the bulk of existing projects.

The data itself is evolving at a faster rate than attitudes towards it. Customer data is growing faster than any other domain and email the fastest growing data source. However, it is the growth of unstructured data which is posing some of the biggest challenges to those surveyed. Approximately one third of respondents believed that unstructured data would constitute more than 70 percent of all data processed within the next five years. Fifteen percent were already there. Finding ways to integrate structured and unstructured data was a key part of the immediate challenge facing our respondents in terms of their data solutions.

Furthermore, panellists indicated that although they were storing ever-increasing quantities of data, that data was not necessarily of sufficient quality to allow analysis of the level required. Meaningful analysis can only be derived from good-quality, integrated data. Getting to this point is the immediate challenge facing businesses.

However, none of this means that organisations cannot be making a start on big data. If businesses start small, ensure a phased approach and a board level sponsor then organisations can start to successfully make the transition to the analysis rich data environment that they aspire to be.

It makes sense to download a free community Hadoop distribution from a recognised vendor so you can be sure of quality, security, performance and continued support. Once that is up and running and real business data problems are being tackled, the task of persuading business decision makers to take big data seriously will become much more achievable.

## **About the sponsor, Intel**

Intel (*NASDAQ: INTC*) is a world leader in computing innovation. The company designs and builds the essential technologies that serve as the foundation for the world's computing devices. Additional information about Intel is available at [newsroom.intel.com](http://newsroom.intel.com) and [blogs.intel.com](http://blogs.intel.com).

