Big Data Revealed
An Information Difference Research Study

December 2013

Sponsored by
helpIT systems  TERADATA
CLEANER DATA. BETTER DECISIONS.  Raising Intelligence
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>4</td>
</tr>
<tr>
<td>BACKGROUND TO THE SURVEY</td>
<td>6</td>
</tr>
<tr>
<td>THE APPROACH</td>
<td>6</td>
</tr>
<tr>
<td>ABOUT THE RESPONDENTS</td>
<td>7</td>
</tr>
<tr>
<td>ADOPTION OF BIG DATA</td>
<td>8</td>
</tr>
<tr>
<td>LEADERSHIP AND FOCUS</td>
<td>10</td>
</tr>
<tr>
<td>KEY BENEFITS AND SUCCESS FACTORS</td>
<td>12</td>
</tr>
<tr>
<td>BIG DATA DOMAINS</td>
<td>15</td>
</tr>
<tr>
<td>COSTS AND SIZE OF BIG DATA</td>
<td>16</td>
</tr>
<tr>
<td>TECHNOLOGIES DEPLOYED</td>
<td>18</td>
</tr>
<tr>
<td>ADDITIONAL FEEDBACK</td>
<td>19</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>21</td>
</tr>
<tr>
<td>ENTERPRISES</td>
<td>21</td>
</tr>
<tr>
<td>VENDORS</td>
<td>22</td>
</tr>
<tr>
<td>ABOUT THE INFORMATION DIFFERENCE</td>
<td>23</td>
</tr>
<tr>
<td>QUESTIONNAIRE</td>
<td>24</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Respondents by Company Revenue</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Respondents by Job Function</td>
<td>7</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Respondents by Industry Sector</td>
<td>8</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Status of Big Data in Organizations</td>
<td>9</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Reasons for Implementing Big Data</td>
<td>9</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Primary Focus for Big Data in Business</td>
<td>10</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Business areas leading Big Data Initiatives</td>
<td>11</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Responsible for Big Data in the Organization</td>
<td>11</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Benefits from Big Data Initiatives</td>
<td>12</td>
</tr>
<tr>
<td>Figure 10</td>
<td>The Main Roadblocks for Big Data Initiatives</td>
<td>13</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Positioning of Big Data and Data Management Initiatives</td>
<td>13</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Data Management Initiatives in Organizations</td>
<td>14</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Critical for Success with Big Data</td>
<td>14</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Is Big Data focused on Data from Web Applications</td>
<td>15</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Key Big Data Domains</td>
<td>15</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Focus on Structured or Unstructured Big Data</td>
<td>16</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Volumes of Big Data Handled</td>
<td>17</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Expenditure on Big Data to Date</td>
<td>17</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Planned Expenditure on Big Data next Year</td>
<td>18</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Software Solutions Deployed (or Planned)</td>
<td>18</td>
</tr>
</tbody>
</table>
EXECUTIVE SUMMARY

There has been a great deal of recent interest in the industry about how to deal with so-called Big Data, which is defined as follows:

**Definition:** Big Data is the term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. (Source: Wikipedia).

Although Big Data has recently received wide press attention, there is still surprisingly little concrete information available regarding the state of Big Data initiatives in business. For example, how many companies have actually implemented Big Data technologies, and in what areas; how much money and effort are organizations investing in it; which parts of the business are driving Big Data investments; what benefits are they seeing; and which data volumes are being handled? This survey, which was sponsored by helpIT and Teradata, aims to address these questions.

The main findings from the survey, based on 178 responses, are summarized below:

- 30% currently have no plans to implement Big Data, 27% currently live with their Big Data initiative, and a further 5% are about to go live. This compares well with the findings from our previous survey of 24%. 24% now plan to go live next year.
- The 27% live implementations splits out regionally as follows: North America (including Canada) 6%, Europe 18%, and the rest of the world 3%. This result flies in the face of conventional wisdom that suggests that North America is well ahead in Big Data.
- 24% want to be able to analyze very large data sets. There does not appear to be a dominant theme as to the reasons for implementing Big Data projects.
- The focus is on improving the quality of decision making (39%), followed by meeting customer needs (a poor second at 19%).
- IT is clearly in the driving seat (33%) with regards to Big Data projects.
- The CIO is generally responsible for and leading Big Data initiatives (31%).
- 13% have identified new business opportunities with their Big Data initiatives. This leads us to conclude that it is still early days when it comes to Big Data.
- The main roadblock to implementation is that it is difficult to present a business case.
- Encouragingly, 45% reported that their Big Data initiative is or will be a part of their wider data management initiatives such as MDM.
- Management understanding and awareness (42%), leadership at the CxO level (18%) and a sound and well-documented business case (9%) are all cited as critical to successful implementation of Big Data initiatives. This suggests there is currently a lack of sufficient buy-in and commitment from the business side.
- Big Data is not mainly focused on analyzing data collected from web-based applications (52%).
- Customer data is the most significant domain for most organizations, with product data second and financial data third. So organizations are not just interested in analyzing Twitter feeds.
- 40% claimed that it is their intention to use Hadoop technologies as the staging area for all their data in future.
- The majority of the data is in the range 1 to 10 terabytes. 5% reported data volumes in the petabyte range, with just 4% in the 200 terabyte to 1 petabyte range. It would seem that Big Data is not necessarily that large.
- 8% have spent over $5 million on their Big Data initiative while 42% have spent less than $1 million, 25% less than $100,000. This further suggests that Big Data is at an early stage in most companies.
• 11% plan to spend more than $5 million while 32% plan to spend less than $1 million in the coming year. 13% will spend less than $100,000.
• The most popular three technologies deployed are Microsoft, Oracle, and Teradata, but there is no dominant technology reported by the survey respondents.
• Respondent feedback indicates that there is a deal of scepticism about the need for business case, viability, and practical implementation of Big Data.
BACKGROUND TO THE SURVEY

There has been a great deal of recent interest in the industry about how to deal with so-called Big Data, which is defined as follows:

**Definition:** Big Data is the term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. (Source: Wikipedia).

Although Big Data has recently received wide press attention, there is still surprisingly little concrete information available regarding the state of Big Data initiatives in business. We believe the time is ripe to survey this area in depth. In particular, we want to put hard numbers around some questions including:

- How many companies have actually implemented Big Data technologies, and in what areas?
- How much money and effort are organizations investing in it?
- Which parts of the business are driving Big Data investments?
- What benefits are they seeing?
- Which data volumes are being handled?

THE APPROACH

The Information Difference survey, “Big Data Revealed”, was conducted over the Internet during the period October to November 2013. The participants were selected by email invitations originating directly from The Information Difference. Participation was also possible via a link from The Information Difference Ltd. website and via links on LinkedIn.

The survey was targeted at senior business and IT leaders worldwide, drawn from larger organizations (with revenues greater than US $1 billion annually).

The participants were provided with the following information prior to completing the survey:

*There has been a great deal of recent interest in the industry about how to deal with so-called Big Data, which is defined as follows:* “Big Data is the term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. (Source: Wikipedia)”

*Despite the plethora of press attention, there is still surprisingly little concrete information available regarding the state of Big Data initiatives in business. At The Information Difference, we consider the time ripe to survey this area.*

*All information provided will be used in aggregate form only and will be kept strictly confidential. The survey has only 20 questions on the topic and should not take more than 10 minutes to complete. In return for a fully completed survey, you will receive a free summary of the analysis of the survey results. Additionally, your name will be entered in a prize draw and the first five winners will receive a free vendor profile (worth $495) of their choice. We will also make a $2 contribution to the Red Cross for each fully completed survey.*

The questionnaire is appended in the section headed Questionnaire.
ABOUT THE RESPONDENTS

A total of 178 respondents worldwide took the survey. 34% were from North America (including Canada), 43% from Europe and the remainder (22%) from the rest of the world.

52% of the respondents were from larger organizations with annual revenues in excess of US $1 billion. Some 21% were from organizations whose annual revenue last year was greater than US$ 10 billion. 48% were from companies with annual revenues below US $1 billion. This represents a balanced mix of both larger and small organizations. The detailed breakdown is shown as Figure 1.

Some 42% of the respondents were drawn from a business background with the majority having an IT role (58%). This likely reflects the current focus of Big Data in the media towards the IT community. 22% had job titles at the Director level or above and 27% had the title of Enterprise Architect. The details are set out in Figure 2.

The largest level of participation was from the banking, insurance, and financial services industry (22%), perhaps, supporting the view that the financial sector is seeking new business opportunities to help it emerge from the financial crisis. The financial sector was the largest vertical of the business respondents with retail and pharmaceuticals/biotech/healthcare both coming in at 9%.

In contrast to our previous survey1 the professional services and computing sectors (32%) were only represented at a combined level of 10%. This earlier survey showed that the main focus and interest

---

1 “Does Big Data mean Big MDM?”, Information Difference Research Report, November 2012
appeared to come from the IT sector. The present results would indicate a shift of interest to the business in the form of the financial sector.

Others included E-Commerce and Recruitment (7%). The remainder represents a wide range of industry sectors. The full results are shown in Figure 3.

![Figure 3 – Respondents by Industry Sector](image)

The analysis of the results from the survey is presented below. The questions referred to in the text are indicated as [Qn] and are set out in full in the appendix headed “Questionnaire”.

Apart from the exceptions mentioned in the text, analysis of the results from the survey for regional dependencies—for example, comparisons between Europe and North America—did not yield any statistically significant differences or trends.

**ADOPTION OF BIG DATA**

The burning question was: How many organizations are actually already up and running with live Big Data implementations? We asked respondents [Q1] to share with us whether they had live initiatives or were planning to implement Big Data within the coming year. The results are summarized in Figure 4.

Some 30% told us that they have currently no plans to implement Big Data in their organizations. Surprisingly, 27% reported that they are currently live with their Big Data initiative, and a further 5% are about to go live. This is very similar to the findings from our previous survey (24%) and suggests that there has been relatively little real progress in terms of the proportion of companies implementing it since 2012. A further 24% plan to go live next year. Taken together, given this level of interest, it would seem that Big Data is beginning to be taken seriously by organizations.
Looking in greater detail at the 27% of live implementations, we find that this splits out regionally as follows: North America (including Canada) 6%, Europe 18% and the rest of the world 3%.

![Figure 4 – Status of Big Data in Organizations](image)

This result flies in the face of conventional wisdom that suggests that North America is well ahead. Certainly, anecdotal evidence has been that the USA has been far ahead of Europe in terms of Big Data take-up, especially as things like web marketing analytics and social media were really the pioneers of Hadoop. Given that 34% of respondents were drawn from North America, compared with 43% from Europe, it cannot reasonably be concluded that this difference can be wholly ascribed to the difference in sample sizes. If we turn our attention to the future plans, then here again we see a slightly higher level of take up in Europe (although much less marked than for live implementations). Of a total of 42% either about to go live or planned, 13% were from North America, 16% from Europe, and 12% from the rest of the world. Might it be that Europe has more interest in the subject than is widely perceived, and that vendors would do well to focus on their European customer base?

We then asked respondents to tell us why they were considering or had initiated their Big Data initiative [Q2]. The results are set out in Figure 5.

![Figure 5 – Reasons for Implementing Big Data](image)
Some 24% are clear that they want to be able to analyze very large data sets. While this clearly makes sense there is a wide range of business demands for Big Data. There does not appear to be a dominant theme. Interestingly “Other” is the second largest category (16%) and includes requirements such as:

- To deal with Customer Requests for Research
- To cope with a large increase in customers
- We want to analyze geo-data as part of our risk assessment
- Analyzing genomic and other patient-related data
- To store medical claim 837 data.

A further 37% divided roughly equally between the ability to be able to analyze web logs (13%), financial transactions (12%), and click-stream data (12%). So the broad focus is on dealing with very large data sets in general, with a wide range of specific drivers.

**LEADERSHIP AND FOCUS**

So what is the focus for Big Data initiatives in most organizations and who is leading these initiatives? We first asked about the primary focus for Big Data initiatives in their organizations [Q3]. The results are shown in Figure 6.

![Figure 6 – Primary Focus for Big Data in Business](image)

Interestingly, the main focus is on improving the quality of decision making (39%), with meeting customer needs a poor second (19%). It is intriguing that cost reduction appears to have been relegated to a lower priority (8%) and, rather surprisingly, management of risk comes in at a mere 5%. Other (which accounts for 13%) included: enhance sales; exploring all (customers, fraud, risk, new business opportunities); identifying "best" prospects for marketing campaigns; increasing the customer experience; etc.

We next asked respondents what areas of their business were leading the Big Data initiative [Q4]? The results are shown in Figure 7.

It is perhaps surprising that IT is so clearly in the driving seat (33%) given the fact that cost reduction (see above) only came in at 8%. Maybe this suggests that IT and business are generally starting to work closer together. Almost a quarter (24%) of the initiatives are being led by Customer Services, which squares up with the previous high focus on “meeting customer needs” (19%).
Other areas rated 16% included: Drilling, Exploring via IT, Fraud and Finance.

Who in the business is actually responsible for Big Data in the organization and for leading the initiatives [Q10]? The responses from the respondents are set out in Figure 8.

In light of the previous result, it is perhaps unsurprising that the CIO is in the driving seat when it comes to responsibility for and the leading of Big Data initiatives (31%). A senior business manager is driving in less than a quarter of organizations (22%). This suggests that many of the initiatives are originating from the IT side of the business and that, in turn, suggests that there may be issues with defining a clear business case.

Other (12%) included: IT Manager, Lead R&D manager, “No one is named or there are many leaders”, “Responsibility is not defined at the time being”, VP Information Systems, chief marketing officer, and “no single person”.

Figure 7 – Business areas leading Big Data Initiatives

Figure 8 – Responsible for Big Data in the Organization
KEY BENEFITS AND SUCCESS FACTORS

What key benefits are organizations seeking and what roadblocks do they identify hampering progress? How does their Big Data initiative relate to other initiatives such as master data management (MDM), data quality, and data governance?

We asked respondents, if they already had a Big Data initiative live, what benefits they had realized to date [Q14]? Their responses are summarized in Figure 9.

![Figure 9 – Benefits from Big Data Initiatives](image)

While some (13%) claimed to have identified new business opportunities, others reported improved operational efficiency (11%), while over half of the respondents didn't know at this time. This strongly suggests that despite the apparent number of “live” implementations, there was as yet no resounding winner for benefits. This leads us to conclude that it still early days.

What are the main roadblocks encountered by organizations [Q9]? Figure 10 shows the results. Respondents were asked to select up to three options, so the results are not additive. The main conclusion appears to be that it is difficult to present a business case. Perhaps we should refer back to Figures 7 and 8 above and view this result in the light of the strong leadership by IT (33%) in many of the initiatives. There is a strong suggestion that lack of clear leadership from the business side may be the root cause of the difficulty in preparing a convincing business case. This is reinforced when we consider that the second most selected option is that there is a lack of clear understanding of the business objectives. One could be forgiven for concluding that it would be good to apply the principle of “more haste less speed” to many of the initiatives. One is forced to the conclusion that in many cases, initiatives have been started without clear business buy-in and business goals. Is the current media hype around Big Data causing organizations to jump too quickly and ill-preparedly into these initiatives?

Is it possible that organizations are jumping into Big Data initiatives too early and before they have paid adequate attention to their data quality and data governance?
We asked respondents to share with us how their current or proposed Big Data initiative was positioned with regard to their other data management initiatives [Q16]. The results are set out in Figure 11.

Very encouragingly, 45% reported that their Big Data initiative is or will be a part of their wider data management initiatives such as MDM. However, this result needs to be interpreted against the
background of which data management initiatives they have already implemented. We asked respondents to tell us about their data management landscape [Q17]. The results are set out in Figure 12.

Figure 12 – Data Management Initiatives in Organizations

It is very positive that more than half of all the organizations have in fact implemented data management across the board. Surprisingly, 54% told us they already have data governance, which is encouraging since this became commonly available much later than master data management (MDM) and data quality. Clearly, we are halfway there and there is still a long way to go in adopting full data management. It is very encouraging that so many organizations see Big Data as integral to their other data management initiatives.

What do organizations believe is critical to success with Big Data initiatives [Q18]? The responses are summarized in Figure 13.

Figure 13 – Critical for Success with Big Data
The responses here are really all focused around securing business buy-in. Management understanding and awareness (42%), leadership at the CxO level (18%), and a sound and well-documented business case (9%) all underline the fact that there is currently a lack of sufficient buy-in and commitment from the business side. The high level of leadership and involvement from IT reported further supports this conclusion. We would suggest that the business would largely appear to be “sitting on the fence” at the moment as far as Big Data is concerned.

**BIG DATA DOMAINS**

So what do organizations believe is Big Data and which data domains are of importance for them? We first asked whether Big Data was mainly focused on analyzing data collected from web-based applications [Q5]. The responses were mixed with a hefty no from 52%, but with a further 36% reporting that it is mostly about web-based application data. Some 9% considered Big Data to be exclusively about web-based application data. The results are shown in Figure 14.

![Figure 14 – Is Big Data focused on Data from Web Applications](image)

More than anything, these results again support the fact that there are clearly multiple drivers for Big Data. We then asked respondents to tell us which data domains were key to their Big Data initiatives [Q6]. The results are shown in Figure 15.

![Figure 15 – Key Big Data Domains](image)
Respondents were asked to select all domains that applied for their business so the results here are not additive. Customer data clearly is the most significant domain for most organizations with product data second and financial data third—interestingly, both after unstructured data, which is clearly gaining in interest. The overall trend is very much in line with what we have found reported for MDM data domains in previous surveys.²

The reference to unstructured data in the previous question is further supported by responses to our next question relating to whether their Big Data initiative was focused on structured or unstructured data [Q7]. The responses are summarized as Figure 16.

![Figure 16 – Focus on Structured or Unstructured Big Data](image)

As can be seen from Figure 16, the response was a resounding “both”. So clearly organizations are not just interested in analyzing Twitter feeds.

There has been increasing interest in the media in the use of Hadoop-type technologies as the “staging area” for all data prior to passing it on to other data stores such as data warehouses or OLAP cubes. We asked our respondents to share with us their views on this use of Hadoop [Q8]. A resounding 40% claimed that it is indeed their intention to use Hadoop technologies as the staging area for all their data in future (prior to loading it into data warehouses or OLAP cubes). Less than a third do not plan to do so, and about one-third doesn’t know at this stage. This is a key result for the conventional ETL vendors, who need either to adapt their technologies or embrace Hadoop in the near future. Otherwise, we suspect they will face increasing difficulties.

## COSTS AND SIZE OF BIG DATA

So how big is Big Data and what are organizations spending in this area? We firstly asked respondents to tell us about the maximum volume of data they are currently handling (or planning to handle) in their Big Data initiative [Q11]. The feedback from the respondents is set out in Figure 17.

From the results reported, it is clear that the majority of the data is in the range 1 to 10 terabytes. Only 5% are reporting data volumes in the petabyte range, with only 4% in the 200 terabyte to 1 petabyte range. Perhaps we should be referring to “quite big data” rather than Big Data. It is surprising that the most popular result lies in the 1 to 10 terabyte range, suggesting that complexity rather than sheer size may be the key driver.

We then turned our attention to what organizations were spending or planning to spend on their Big Data initiatives.

What is Big Data costing? We asked organizations what they have spent to date on their Big Data initiatives [Q13]. The expenditure range is set out in Figure 18.

We also asked what they plan to spend on Big Data in the coming year [Q12]? The expenditure range is set out in Figure 19. Only some 8% have spent over $5 million, while 42% have spent less than $1 million, 25% less than $100,000. This suggests that it is still very early for most Big Data initiatives with relatively modest outlays to date. Roughly a third either don't know or have spent nothing.

Turning to their planned expenditure over the next twelve months, 11% plan to spend more than $5 million while 32% plan to spend less than $1 million. 13% will spend less than $100,000. While this suggests a slight increase in planned expenditure over the coming twelve months, it is not a major commitment. This leads us to conclude that Big Data is still in the early days of being adopted, with many organizations perhaps testing the temperature of the water with relatively limited expenditure.

It seems fair to conclude that Big Data has some way to go before it reaches even the first stages of maturity.
What technologies are being deployed by organizations for their Big Data initiatives? We asked respondents to indicate which of the current range of software tools for Big Data they either had deployed or planned to use [Q15]. The results are presented as Figure 20.

**Figure 20 – Software Solutions Deployed (or Planned)**
The respondents were asked to select all that applied, so the results are not additive and do not represent market share.

The most popular three appear to be Microsoft, Oracle, and Teradata. Most of the remainder have been deployed or planned by a smaller number of organizations. It is interesting to note that key vendors such as Hortonworks, MapR, and Jaspersoft do not figure high on this list. Perhaps this is because in the first instance, organizations have turned to their current software suppliers to provide Big Data solutions. We see this as a likely future trend with business expecting their house vendor, be it SAP, IBM, Microsoft, Informatica, or Oracle, to provide a Big Data solution.

Other software solutions mentioned included: AWS Redshift, Cassandra, Causata, Infobright, QlikView, and SAS.

**ADDITIONAL FEEDBACK**

Finally, we asked respondents to provide any additional views or comments that they had in regard to the role of Big Data in their organization [Q19]. Their extensive feedback included:

- As of now it (our Big Data initiative) is being used as a huge archival system for all transaction data for further analysis.
- Big Data is not only about having a Hadoop cluster!
- Big Data seems to be buzz word only for us—needs to be more concrete.
- Big data is and will be more important.
- Data governance and good quality and accurate data is key for decision-making processes.
- I don’t think Big Data will help organizations in the analytics area.
- It is all a question of further trial and error...
- It will have a key role in our Information management.
- It’s a solution looking for a problem in many cases.
- It is a crucial part of financial reporting and controlling.
- Lots of moving parts with evolving use cases
- Sponsorship needed.
- The ability to separate the “Big Noise” from “Big Data” will be crucial for the success.
- Too soon to know what the proper approach is.
- We need easy instructions.
- We are still in the experimental stage, learning the technologies.
- We will develop our own tools.
- Big Data is a paradigm shift in IT thinking and business needs to be continuously involved as implementation partner.
- Organizations with immature BI/DG will eventually make a huge, costly erroneous decision that “no one could have predicted” a la Taleb’s Black Swan warning.
- We not have a Big Data initiative internal to the organization, but are creating offerings for our customers.
- We simply haven’t identified a use case yet—it isn’t needed, and our current processes suffice.
- Difficult to find a good business use case for it and then be able to demonstrate the value in a cost-effective way.
- The key challenge will be to integrate Big Data with existing internal data sources. So data integration and data quality are critical enablers.
• Our Big Data use case exists, but it is an evolutionary step and requires commitment from end customers to implement.
• Understand the opportunities, have a long term strategy, future proof the organization, gain executive support, create business case and justification, start with low hanging fruit like operational efficiencies.
• As a business user, questions regarding data size are completely irrelevant. I'm interested in capabilities—the tech choices are not of interest.
• We are trying to work with our marketing area and Medical Informatics area to find an appropriate business case for Big Data. We are going to look at web logs and analyze large data sets to do some healthcare predictive modeling.
• The key goal is predictive analytics applied to machine data, maintenance event data and operational data in order to understand and predict large fleet performance and optimize the maintenance regime.
• At this moment, there are very few people who can really analyze the Big Data paradigm without being carried away by the hype. It is crucial to know how it's going to help your particular business case instead of trying to be on the latest bandwagon.
• Before you start, you need to understand how you will manage Big Data, and the scalability of your solution.
• Big data in a healthcare setting is maybe of a different proportion than Big Data as it's defined in the world of telco, social media, etc., but it's really big.
• Big Data is an extension of our BI/Data Warehousing program. Big Data extends existing options by opening up a depth of data. Big Data tools also open the door for new types of analytical techniques.
• We are exploring Big Data, and trying to understand how it may be applied and what benefits it would provide for our business.
• BD is not mature, yet some bona fide successes are demonstrated—a difficult combination to sell. Further, we still have ways to go with classic RDBMS data: DQ, DG, etc.

It is clear from the respondent feedback above that there is a deal of skepticism about the need for, business case, viability and practical implementation of Big Data.
CONCLUSIONS

Key conclusions and recommendations resulting from the survey analysis are summarized below. These have been split into two groups: those of direct relevance to enterprises and organizations considering or in the process of implementing (or who have already implemented) Big Data initiatives, and those relating to the software vendors and systems integrators (SIs).

Enterprises

- 30% told us that they currently have no plans to implement Big Data. 27% reported that they are currently live with their Big Data initiative, and a further 5% are about to go live. This compares well with the findings from our previous survey (24%), and suggests that there has been relatively little progress since 2012. A further 24% plan to go live next year.
- The 27% live implementations split out regionally as follows: North America (including Canada) 6%, Europe 18%, and the rest of the world 3%. This result flies in the face of conventional wisdom that suggests that North America is well ahead.
- 24% want to be able to analyze very large data sets. Clearly, there is a wide range of business demands for Big Data. There does not appear to be a dominant theme.
- The main focus is on improving the quality of decision making (39%), with meeting customer needs a poor second (19%).
- IT is clearly in the driving seat (33%). Almost a quarter (24%) of the initiatives are being led by Customer Services.
- Given the previous result it is perhaps unsurprising that the CIO is in the driving seat as responsible for and leading Big Data initiatives (31%).
- Some 13% claimed to have identified new business opportunities with their Big Data initiatives. Others reported improved operational efficiency (11%) while over half of the respondents didn’t know at this time. This strongly suggests that despite the apparent number of “live” implementations, there was as yet no resounding winner for benefits. This leads us to conclude that it still early days.
- The main roadblock to implementation appears to be that it is difficult to present a business case.
- Encouragingly, 45% reported that their Big Data initiative is or will be a part of their wider data management initiatives such as MDM. More than half of all the organizations have in fact implemented data management across the board (MDM, data quality and data governance).
- Management understanding and awareness (42%), leadership at the CxO level (18%) and a sound and well-documented business case (9%) are all cited as critical to successful implementation of Big Data initiatives. This suggests there is currently a lack of sufficient buy in and commitment from the business side.
- Big Data is not mainly focused on analyzing data collected from web-based applications.
- Customer data is the most significant domain for most organizations with product data second and financial data third, both after unstructured data which is clearly gaining in interest. So clearly organizations are not just interested in analyzing Twitter feeds.
- 40% claimed that it is their intention to use Hadoop technologies as the staging area for all their data in future (prior to loading it into data warehouses or OLAP cubes), so the possible impact on enterprise ETL strategy needs to be considered.
The majority of the data is in the range 1 to 10 terabytes. Only 5% reported data volumes in the petabyte range with only 4% in the 200 terabyte to 1 petabyte range. Big Data may be as much about complexity as sheer size.

Some 8% have spent over $5 million on their Big Data initiative, while 42% have spent less than $1 million, 25% less than $100,000. This suggests that it is very still early when it comes to Big Data initiatives with relatively modest outlays to date.

Over the next twelve months, 11% plan to spend more than $5 million, while 32% plan to spend less than $1 million. 13% will spend less than $100,000.

The most popular three technologies deployed by survey respondents are Microsoft, Oracle, and Teradata, but there is no dominant vendor at this stage.

Respondent feedback indicates that there is a deal of skepticism about the need for, business case, viability and practical implementation of Big Data.

### Vendors

- Roughly one-third of organizations currently have no plans for Big Data, while 24% plan to go live in the coming year. This represents a significant opportunity for vendors for expansion.
- Our findings suggest that take-up is faster in Europe that North America. This offers a double challenge for vendors to both continue support and growth in Europe while encouraging growth in North America.
- 24% want to analyze very large data sets, but we note a wide range of business requirements for Big Data. Vendors need to focus on developing their strategies and product offerings to be in a strong position to respond to these requirements. We suggest that vendors need to adopt much more solution-focused approaches for business.
- Only 13% claimed to have identified new business opportunities while over half were unclear as to the benefits at this time. The main roadblock was the difficulty of presenting a business case. Vendors need to focus more on developing end-user case studies that will help organizations to better understand and explain the potential benefits. There appear to be very few case studies currently available.
- Management understanding and awareness (42%), leadership at the CxO level (18%), and a sound and well-documented business case (9%) are all cited as critical to successful implementation of Big Data initiatives. Vendors clearly need to address this with better-targeted education at the senior management level.
- Customer data is the most significant Big Data domain for most organizations with product data second and financial data third, both after unstructured data which is clearly gaining in interest. Organizations are clearly not just interested in analyzing Twitter feeds. Vendors need to ensure that they are able to provide solutions that address the customer domain.
- 40% claimed that it is their intention to use Hadoop technologies as the staging area for all their data in future (prior to loading it into data warehouses or OLAP cubes). This poses a significant risk for those conventional ETL vendors and they would do well to address this urgently.
- Planned expenditure over the next twelve months is limited. 11% plan to spend more than $5 million while 32% plan to spend less than $1 million.
- The most popular technologies deployed are Microsoft, Oracle, and Teradata, though with no dominant vendor emerging at this stage. Most of the specialist Big Data vendors do not appear high on the list of technologies deployed. This suggests that organizations are looking to their main software suppliers to provide technologies to deal with Big Data requirements rather than selecting best of breed. Larger software vendors need to ensure that they have viable offerings
for their client base and niche or smaller vendors need to seek ways to offer more solution oriented products.

- Feedback from respondents indicates that there is a deal of skepticism about the need for, business case, viability and practical implementation of Big Data. Vendors need to address this concern by providing more case studies and education on the benefits and opportunities for Big Data.

ABOUT THE INFORMATION DIFFERENCE

The Information Difference is an analyst firm focusing primarily on master data management (MDM), data quality, and data governance. Our founders are pioneers who helped shape the MDM industry with in-depth global project experience. We offer detailed analysis of these industries, in-depth profiles of the MDM and data quality vendors, assessments of the marketplace and white papers discussing key issues and best practice. Additionally, we can offer advice on strategy, vendor selection and best practice in these areas. We carry out primary market research and can help you with MDM project justification, building the business case and return on investment.
QUESTIONNAIRE

The full questionnaire used in the survey is included below. The navigation logic is not shown in the interests of clarity.

Big Data Revealed

There has been a great deal of recent interest in the industry about how to deal with so-called Big Data, which is defined as follows: “Big Data is the term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. (Source: Wikipedia)”

Despite the plethora of press attention there is still surprisingly little concrete information available regarding the state of Big Data initiatives in business. At The Information Difference we consider the time is ripe to survey this area.

All information provided will be used in aggregate form only and will be kept strictly confidential. The survey has only 20 questions on the topic and should not take more than 10 minutes to complete. In return for a fully completed survey you will receive a free summary of the analysis of the survey results. Additionally your name will be entered in a prize draw and the first five winners will receive a free vendor profile (worth $495) of their choice. We will also make a $2 contribution to the Red Cross for each fully completed survey.

Please note that questions marked with an asterisk (*) are mandatory.

1) Have you a Big Data initiative or are you planning such an initiative in the next year?*
   - Currently live
   - About to go live
   - Planned for the current year
   - Planned for next year
   - No current plans
   - Don't know

2) Why are you considering/initiating a Big Data initiative?*
   - We want to be able to analyze social media activity (including sentiment analysis)
   - We want to be able to analyze click-stream data
   - We want to be able to analyze web logs
   - We want to be able to analyze financial transactions
   - We want to be able to analyze video data
   - We want to be able to analyze sensor/machine data from equipment
We want to be able to analyze very large data sets

Other (Please specify):

Don't know

3) What is the primary focus of Big Data in your organization (e.g.: customers, risk, new opportunities, ...)?*

- Management of risk/fraud
- Meeting customer needs
- Seeking new product and/or business opportunities
- Improving the quality of our decision making
- Identifying/reducing costs
- Other (Please specify):

Don't know

4) What areas of your business are leading your Big Data initiatives?*

- Information Technology
- Sales and Marketing
- Customer Services and/or Customer Relationship Management
- Online operations (e.g.: eBusiness, eCommerce, ...)
- Risk Management
- Fraud Detection
- Operations and Supply Chain
- New Product Development / R&D
- Other (Please specify):

Don't know

5) Is Big Data mainly about analyzing data collected from web-based applications?*

- Yes, exclusively
- Yes, mostly
- No
- Don't know
6) Which data domains are key to your Big Data initiative?

[Please select all that apply]*

- Name and address data (of customers or suppliers)
- Product data
- Financial data
- Supply chain data
- Human resources (HR) data
- Customer & Prospect Data
- Fraud and Risk Management Data
- Customer Service Data
- Unstructured data
- Telephone call logs
- Don't know
- Other data (Please specify)

7) Is your Big Data initiative focused on unstructured or structured data or both?*

Data can be designated as structured or unstructured data for classification within an organization. The term structured data refers to data that is identifiable because it is organized in a structure, such as records in a database. In contrast, unstructured data has no identifiable structure, such as images, video, audio, and email.

(Source: webopedia)

- Structured Data
- Unstructured Data
- Both
- Don't know

8) Are you planning to use Hadoop as the 'staging area' for all your data prior to passing it on to other data stores such as data warehouses or OLAP cubes?*

- Yes
- No
- Don't know

9) What are the main roadblocks you have encountered (or expect to encounter) with Big Data initiatives?

[Please select up to three]*

- It’s very difficult to present a business case
- Lack of clear understanding of the business objectives for Big Data
It’s difficult to identify where to find help
We do not have the right skill sets
Management does not see this as an imperative
It would be too expensive
Unrealistic expectations are often set
No one is prepared to lead the initiative
No one in the business seems to care
Can’t find suitable technology
Can’t find data engineers with the necessary talent/expertise
Other data management issues take precedence
Don’t know
Other (Please specify)

10) Who is responsible for Big Data in your organization?*

- Chief Executive Officer
- Chief Finance Officer
- Chief Operations Officer
- Chief Information Officer
- Enterprise (or Chief) Architect
- Senior Business Manager
- Other (Please specify):

- Don’t know

11) What is the maximum volume of data that you are currently handling/plan to handle in your Big Data initiative?*

- Less than 1 terabyte
- 1 to 10 terabytes
- 10 to 50 terabytes
- 50 to 100 terabytes
- 100 to 200 terabytes
- 200 to 500 terabytes
- 500 to 1000 terabytes
- Greater than 1 petabyte (1000 terabytes)
- Don’t know
12) What do you plan to spend on Big Data in the next year?*
- More than $10 million
- $5 million to $10 million
- $1 million to $5 million
- $500,000 to $1 million
- $100,000 to $500,000
- Less than $100,000
- We have no plans
- Don't know

13) What have you spent on Big Data to date?*
- More than $10 million
- $5 million to $10 million
- $1 million to $5 million
- $500,000 to $1 million
- $100,000 to $500,000
- Less than $100,000
- We have no plans
- Don't know

14) If you already have a Big Data initiative, what benefits have you realized to date?*
- More reliable decision making
- Identified new product opportunities
- Identified new business opportunities
- Improved management of risk (reduced risk)
- Improved customer service
- Improved operational efficiency
- Other (Please specify):
- Don't know

15) Which Big Data products/tools are you considering/using for your Big Data initiative?
[Please check all that apply]*
- 1010data
- 10Gen (Mongo DB)
- Actian (was Ingres, ParAccel)
- Amazon (Dynamo DB)
- Amazon (Elastic MapReduce, Redshift)
- Attensity
- Attivio
- Attunity
- Birst
- Cambridge Semantics
- Clarabridge
- Cloudant
- Cloudera
- Composite
- Couchbase
- Data Stax
- Datameer
- Datasift
- Denodo
- Guavas
- Hadapt
- Hortonworks
- HP
- HPCC
- IBM
- Impetus
- Informatica
- Jaspersoft
- Karmasphere
- Kognitio
- MapR
- Microsoft
- Neo
- NuoDB
- Oracle
- Palantir
- Pentaho
16) How are you positioning your Big Data initiative with respect to your other data management initiatives?*

- Big Data is an integral part of data management initiatives such as Master Data Management, Data Quality and Data Governance
- Big Data is a totally separate area from our data management initiatives
- Big Data is more closely aligned with our data quality initiative
- Big Data is more closely aligned with our Master Data Management program
- Don't know
- Other (Please specify):

17) Which of the following have you already implemented?*

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Planned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master Data Management (MDM)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18) Which of these do you believe is/will be crucial to success with Big Data, if any?*

- Leadership at the CxO level
- Management understanding and awareness of the opportunities presented by Big Data
- A sound well documented business case
- Availability and access to skilled resources
- Identifying the best tools to use
- Beginning with clean, high quality data
19) Please enter below any additional views/comments which you may have in regard to the role of Big Data in your organization.


20) What was your company's total revenue last year?*

- More than $50 billion
- $1 billion to $10 billion
- $500 million to $1 billion
- $100 million to $500 million
- Less than $100 million
- $10 billion to $50 billion

21) Please select the main industry in which your company operates.*

- Aerospace & Defense
- Agriculture
- Banking/Insurance/Financial Services
- Chemicals/Energy/Utilities
- Computing (Hardware and/or Software)
- Construction
- Education/Training
- Government-Federal/State/Local
- Leisure/Travel/Hospitality
- Manufacturing
- Media/Publishing/Entertainment
- Metals & Mining
- Non-Profit/Charitable
- Pharmaceuticals/Biotech/Healthcare
- Professional Services/Consulting
- Real Estate
22) Which of the following best describes your title or role in your company?*
- CxO, SVP or other Executive Role
- VP, General Manager, Director
- CIO or VP of Information Technology
- Enterprise Architect or Chief Architect
- Other Business Title
- Other IT Title

23) Are you willing to take part in a brief, confidential discussion on this topic with an Information Difference analyst?*
- Yes
- No

24) Would you be willing to share your contact information with our research sponsors in order to learn more about their products?*
- Yes - helpIT
- Yes - Teradata
- No

25) Please provide your brief contact details:
First Name*:

Last Name*:

Company Name*:

Email Address*:

26) Please select your country from the drop down list.*