



Is the Data Warehouse Dead?

An Information Difference Research Study

January 2015

Sponsored by



TABLE OF CONTENTS

EXECUTIVE SUMMARY	4
BACKGROUND TO THE SURVEY	5
THE APPROACH.....	5
ABOUT THE RESPONDENTS	6
CURRENT STATE OF DATA WAREHOUSING IN ORGANIZATIONS	8
ADOPTION OF BIG DATA SOLUTIONS BY ORGANIZATIONS	12
THE DEMISE OF THE DATA WAREHOUSE?	15
CONCLUSIONS.....	18
ENTERPRISES	18
VENDORS.....	19
ABOUT THE INFORMATION DIFFERENCE	20
QUESTIONNAIRE	21

LIST OF FIGURES

<i>Figure 1 - Respondents by Revenue.....</i>	<i>6</i>
<i>Figure 2 - Respondents by Job Function.....</i>	<i>7</i>
<i>Figure 3 - Respondents by Industry Sector</i>	<i>7</i>
<i>Figure 4 - Proportion of Organizations with Data Warehouses</i>	<i>8</i>
<i>Figure 5 - Number of active Data Warehouses.....</i>	<i>9</i>
<i>Figure 6 - Size of active Data Warehouses.....</i>	<i>9</i>
<i>Figure 7 - Assessment of Running Costs of Data Warehouse</i>	<i>10</i>
<i>Figure 8 - Overall Assessment of Data Warehouse(s)</i>	<i>11</i>
<i>Figure 9 - Ability to handle Unstructured Data</i>	<i>11</i>
<i>Figure 10 - How important is Big Data for your Organization?</i>	<i>12</i>
<i>Figure 11 - Have you an active Big Data initiative?</i>	<i>12</i>
<i>Figure 12 - Size of Big Data Implementations.....</i>	<i>13</i>
<i>Figure 13 - Ability to handle Unstructured Data</i>	<i>14</i>
<i>Figure 14 - How happy are you with your Big Data implementation?.....</i>	<i>14</i>
<i>Figure 15 - Will Big Data solutions replace the data warehouse?</i>	<i>15</i>

EXECUTIVE SUMMARY

Big Data has generated much interest and attention in the media of late. Indeed, several authors have recently raised the question of whether Big Data approaches, such as Hadoop, will pronounce the death sentence on the conventional data warehouse.

"A data warehouse (DW, DWH), or an enterprise data warehouse (EDW), is a system used for reporting and data analysis. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Data warehouses store current and historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons." (Source: Wikipedia)

"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)

In this survey we investigate the current state of the data warehouse and examine its recent challenger in the form of Big Data solutions as an alternative. Is the new technology really complementary or is the reign of the data warehouse nearing an end? The main findings from the survey, based on responses from more than 90 organizations worldwide, are summarized below:

- 88% of organizations have at least one data warehouse. Most (45%) have between 2 and 5 warehouses, with all the associated issues of maintaining data consistency.
- Only 23% have managed to get down to just a single data warehouse.
- 34% have less than 10 terabytes of data suggesting that Big Data may well not be for everyone.
- Most organizations (41%) have data warehouse sizes in the range 1 to 50 terabytes, though 4% have to deal with warehouses larger than a petabyte.
- The costs of support and maintenance (expressed as full time equivalent staff or FTE) varied significantly, with a mean of 19 FTE and a median of 7 FTE. Clearly data warehouse maintenance is a non-trivial burden.
- Some 27% of organizations expressed unhappiness with the costs of their data warehouse.
- In general, organizations are happy (53%) with their data warehouse(s).
- However, 55% were unimpressed with the ability of their data warehouse to handle unstructured data (e.g.: web logs, text, sensor data, ..).
- Two thirds (64%) of organizations consider Big Data to be important for their business.
- Around one fifth (22%) of organizations already have a live Big Data initiative while a further 11% are about to go live.
- Some 37% of the current Big Data implementations are less than 100 terabytes, with 30% being less than 50 terabytes, but 11% over 500 TB.
- Amongst those with live Big Data implementations there is a clear view (43%) that Big Data technologies are much better at handling unstructured data.
- Of those with a Big Data implementation, 3 times as many are happy with it as unhappy..
- The general view (43%) is that data warehousing and Big Data technologies are currently, and will remain, complementary. The data warehouse is certainly not dead.
- Only 2% claimed that Big Data technologies would replace data warehousing.

Based upon the broader feedback from the survey, we suggest that the technologies need to be closely aligned with conventional data warehousing (together with MDM) ensuring data consistency for subsequent use by Big Data technologies. This view is highlighted in particular by one quote from the survey: *"I suspect that "Big Data" is a way to THINK that you are obtaining good data while avoiding the hard work of understanding and designing data models."* This is certainly a potential pitfall.

BACKGROUND TO THE SURVEY

Big Data has generated much interest and attention in the media of late. Indeed, several authors have recently raised the question of whether Big Data approaches, such as Hadoop, will pronounce the death sentence on the conventional data warehouse.

"A data warehouse (DW, DWH), or an enterprise data warehouse (EDW), is a system used for reporting and data analysis. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Data warehouses store current and historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons." (Source: Wikipedia)

"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)

In this survey we investigate the current state of the data warehouse and examine its recent challenger in the form of Big Data solutions as an alternative. Is the new technology really complementary or is the reign of the data warehouse nearing an end? Specifically, we will address such questions as:

- How successful are current data warehouses?
- How many warehouses do companies really have deployed today?
- How do end-users perceive Big Data alternatives, and how widely are these really deployed?
- Are companies using Big Data in ways that overlap with current data warehouses?
- Do end-users plan to replace current data warehouse technology eventually?

THE APPROACH

The survey, entitled ***"Is the Data Warehouse Dead?"***, was conducted over the Internet during the period November 2014. 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.

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

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

"A data warehouse (DW, DWH), or an enterprise data warehouse (EDW), is a system used for reporting and data analysis. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Data warehouses store current and historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons." (Source: Wikipedia)

"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)

Big Data has generated much interest and attention in the media of late. Indeed, several authors have recently raised the question of whether Big Data approaches, such as Hadoop, will pronounce

the death sentence on the conventional data warehouse. At The Information Difference we considered it opportune to explore the views of the data warehouse user community.

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 of their choice. We will also make a \$2 contribution to the Red Cross for each fully completed survey.

The full questionnaire is appended in the section headed Questionnaire.

ABOUT THE RESPONDENTS

More than 90 companies and organizations from across the world completed the survey. 44% were from North America (including Canada), 27% from Europe and the remainder (29%) from the rest of the world.

Almost two thirds (62%) of the respondents were from larger organizations with annual revenues in excess of US \$1 billion. Some 13% were from organizations whose annual revenue last year was greater than US\$ 50 billion. 38% were from companies with annual revenues below US \$1 billion. This represents a broad span of both larger and small organizations. The detailed breakdown is shown as Figure 1.

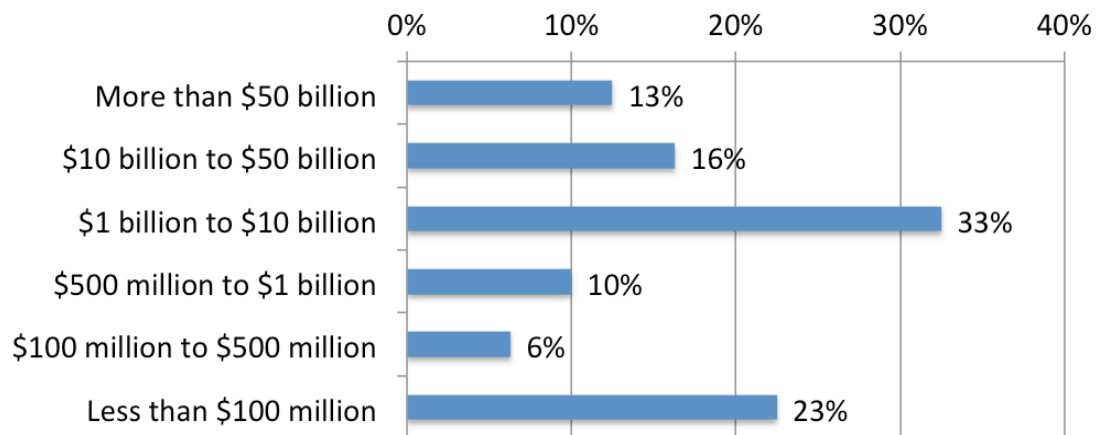


Figure 1 - Respondents by Revenue

Only 19% of the respondents were drawn from a business background with the majority having an IT role (80%). This likely reflects the current focus of Big Data discussions in the media towards the IT community. 20% had job titles at the Director level or above and 35% had the title of Enterprise Architect. The details are set out as Figure 2.

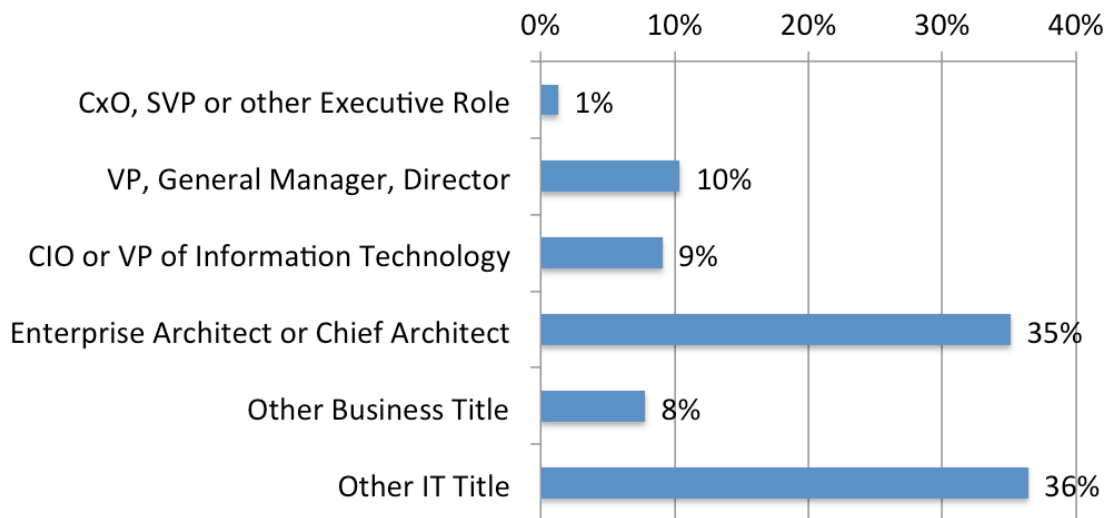


Figure 2 - Respondents by Job Function

The highest level of participation was from the banking, insurance, and financial services industry (26%), perhaps further supporting the notion that the financial sector is seeking to reduce costs and identify new business opportunities to help it emerge from the financial crisis.

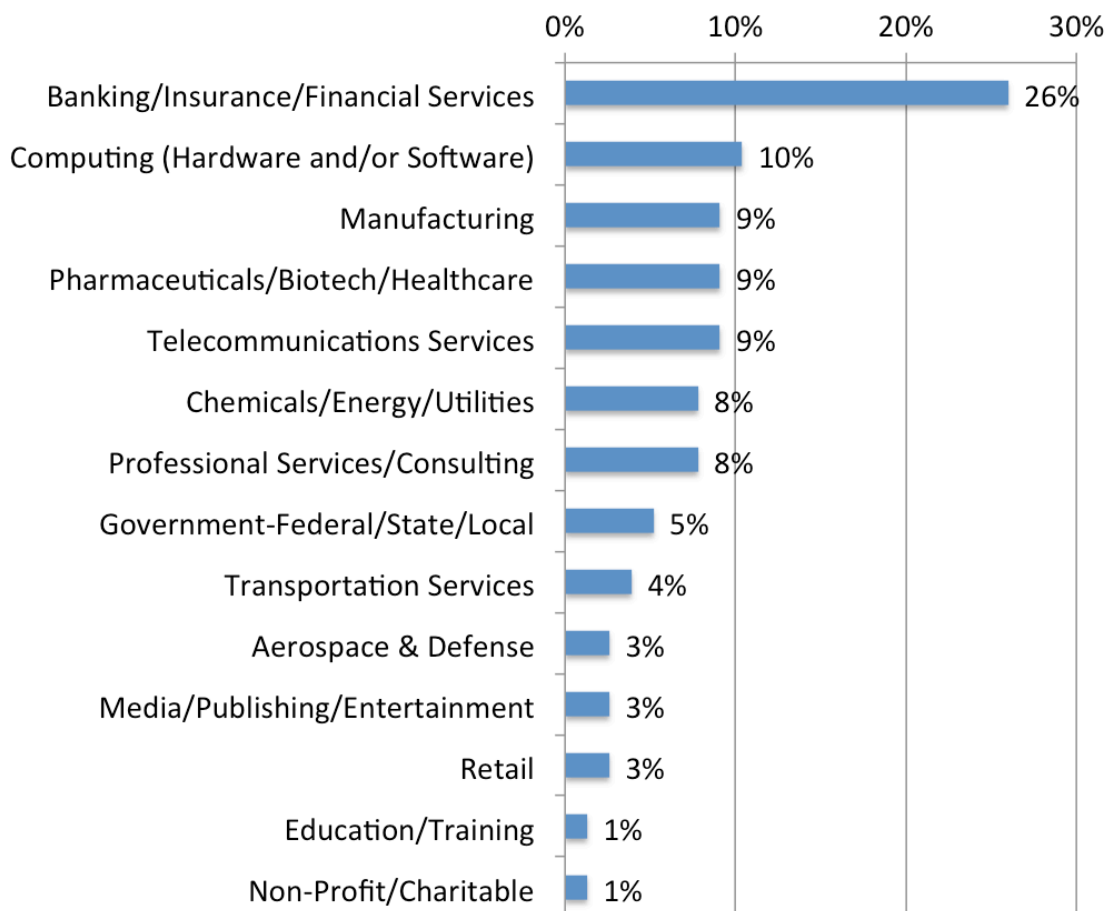


Figure 3 - Respondents by Industry Sector

Computing took second place (10%) reflecting the strong current focus by Big Data software vendors on the IT community. Manufacturing (9%), pharmaceuticals, biotech and healthcare (9%) together with telecommunication services (9%) jointly took joint third place. This is unsurprising since these industries probably have the greatest requirement for Big Data analysis.

The remainder represents a wide range of industry sectors. The full results are shown in Figure 3.

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.

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.

CURRENT STATE OF DATA WAREHOUSING IN ORGANIZATIONS

What is the current position of data warehousing across a wide range of organizations? We first asked respondents to tell us whether they had one or more active data warehouses in their organization [Q1]. Some 88% responded that they had active data warehouses with just 7% claiming to have none. The results are shown as Figure 4.

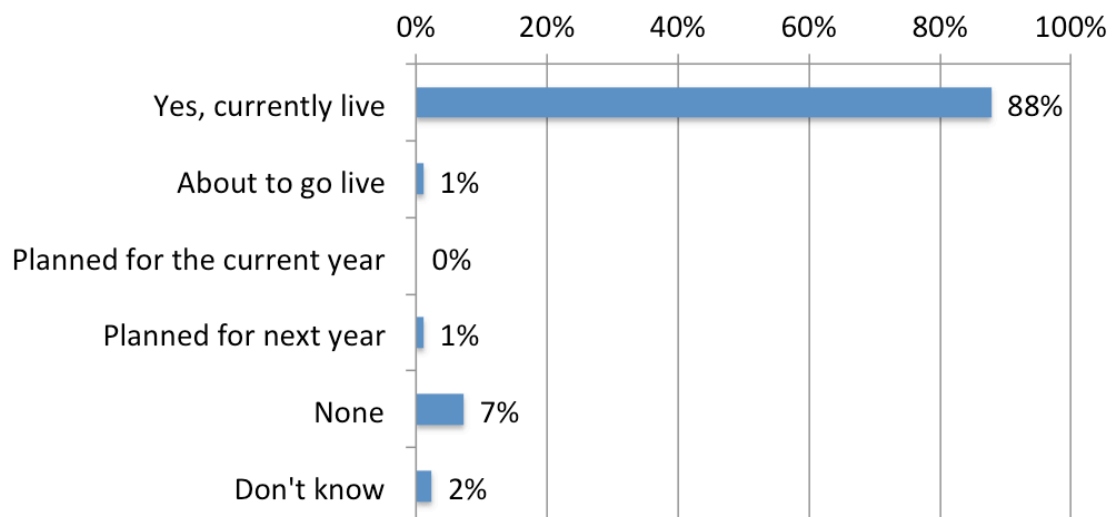


Figure 4 - Proportion of Organizations with Data Warehouses

It is interesting to note that 7% do not have a data warehouse, which is a little puzzling given that the technology is now very mature and all organizations require some form of management reporting. As a follow up question [Q2] we asked respondents to tell us (or estimate) how many separate data warehouses they had across their organization. Most organizations surveyed had between 2 and 5 (45%) data warehouses with 23% reporting that they had just one. The results are summarized in Figure 5.

So around two thirds of organizations have fewer than 5 active data warehouses with just 6% reporting more than 50.

Next we sought to understand how large were their data warehouses and what level of resources was general required (in terms of Full Time Equivalent {FTEs}) to support these.

It is interesting to discover that only 23% have got down to just a single data warehouse. Indeed 10% have more than 20 warehouses, with all the associated costs and complexity issues of consistency that that implies. This is all the more surprising given that focus in recent years in the media on the value of moving to a single data warehouse.

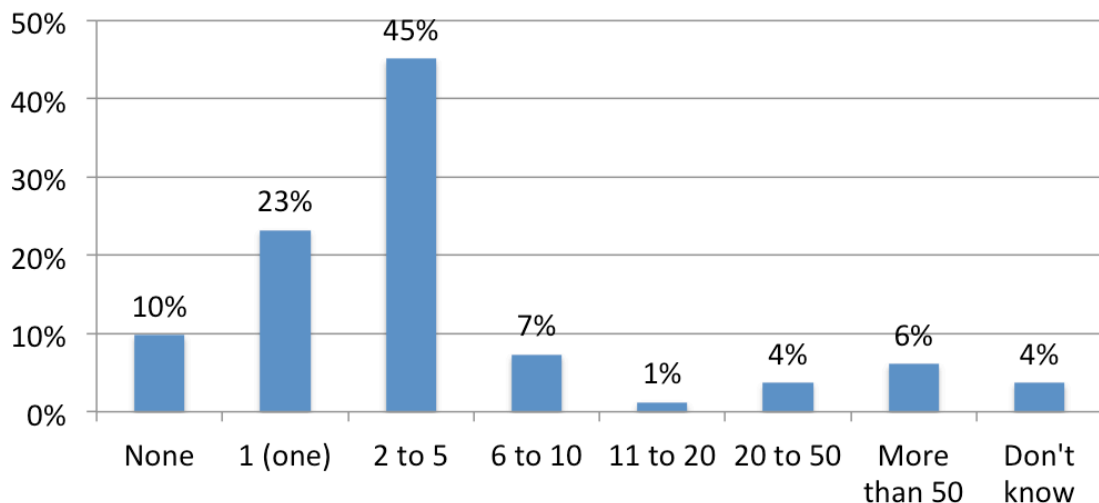


Figure 5 - Number of active Data Warehouses

We asked respondents to provide us with an estimate of the size of their data warehouses [Q5]. The results are set out in Figure 6.

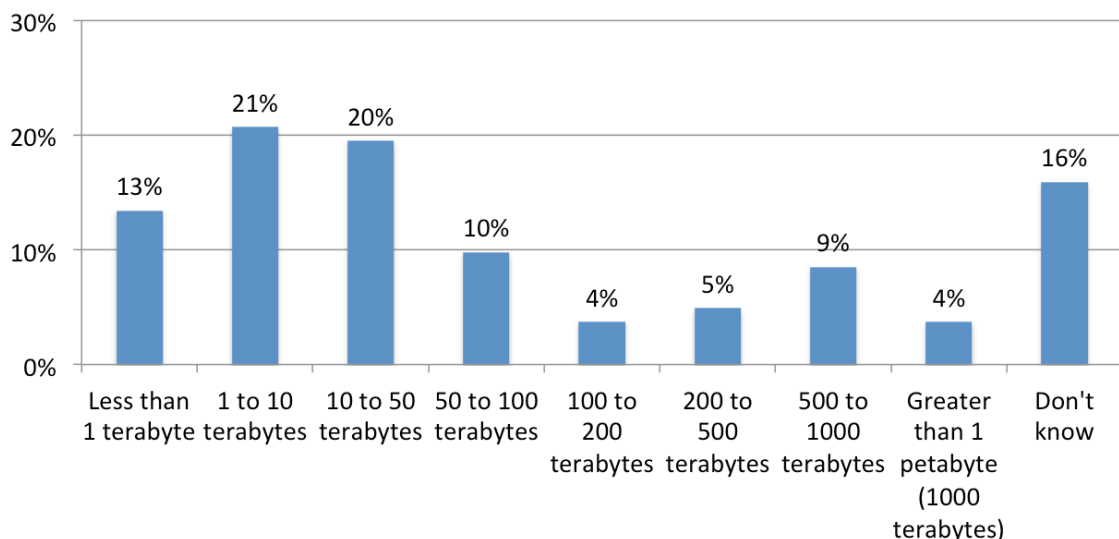


Figure 6 - Size of active Data Warehouses

Interestingly, the majority of data warehouses currently active in organizations are less than 50 terabytes. The result that 34% have less than 10 terabytes suggests that Big Data is certainly not for everyone. Only 4% claimed to have data warehouses larger than 1 petabyte. In general, sizes lie within the range 1 to 50 terabytes for most companies. This is significant since 50 terabytes is not an

especially large database. Some 13% have in excess of 500 terabytes and the range 10 to 100 terabyte appears to be the most common range (30%).

We then asked respondents to share with us the levels of resources that were deployed to maintain and support their data warehousing [Q6]. The mean value was 19 full-time equivalent staff (FTE) with a median of 7 FTE. The range was broad and varied from 100 to 0 FTE. It is worth noting that a median of 7 FTE implies a substantial level of maintenance resource.

The overall picture appears to be that most organizations have relatively small to medium sized data warehouses requiring some 7 FTE for support and maintenance. Relatively few (6%) have very large data warehouses where the focus might indeed be more on Big Data approaches.

What was the overall assessment of respondents of the operational or running costs of their corporate data warehouses [Q4]. The results are given in Figure 7.

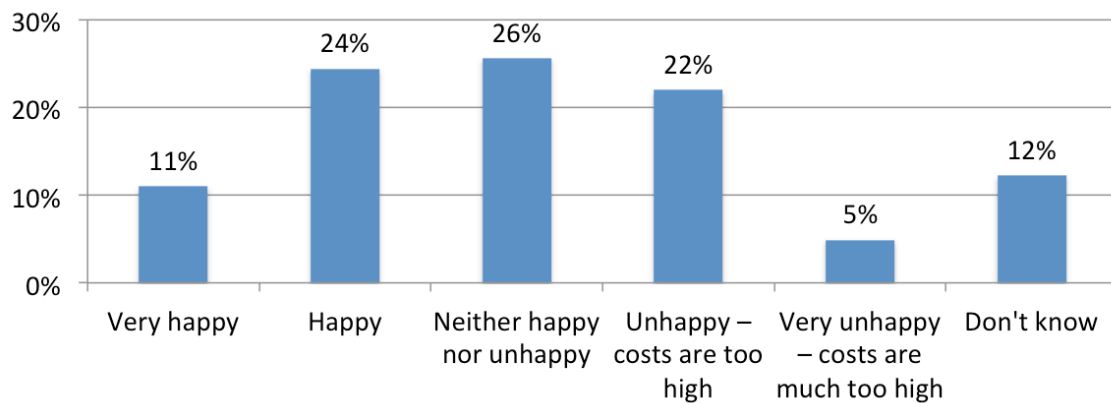


Figure 7 - Assessment of Running Costs of Data Warehouse

About one third (35%) reported that they were at least happy with the running costs of their current active data warehouse. This contrasts with 27% who reported that they are unhappy with the current costs – regarding these as being too high. A further 26% sit in the middle region and are neither happy nor unhappy. It appears that there is some measure of concern over costs associated with running and maintaining their active data warehouses.

So what is their overall assessment of their current data warehouse(s) in terms of meeting the business needs [Q3]? Generally respondents appear to be fairly contented with their current data warehouse. The results are presented as Figure 7.

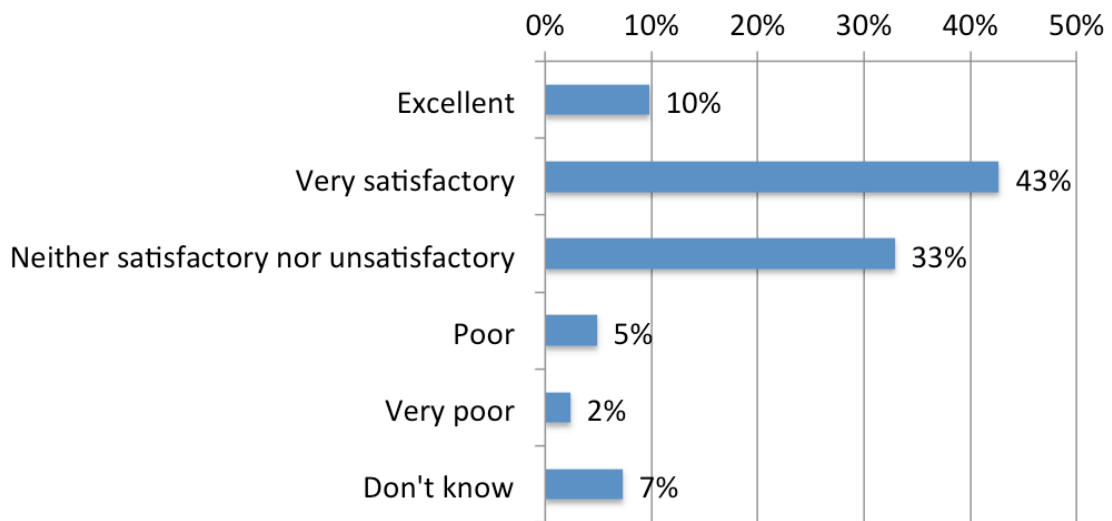


Figure 8 - Overall Assessment of Data Warehouse(s)

So some 53% are generally very happy with their current data warehouse while only 7% reported that they considered it poor. This is interesting when compare with the responses shown in Figure 7 where there is clearly some concern over the level of costs. This may be partly explained by the fact that many organizations still have multiple data warehouses each needing maintenance to ensure consistency.

Finally, in this section, we asked the view of the respondents on the ability of their current data warehouses to handle unstructured data such as text, web logs and sensor data [Q7]. Their views are summarised in Figure 9.

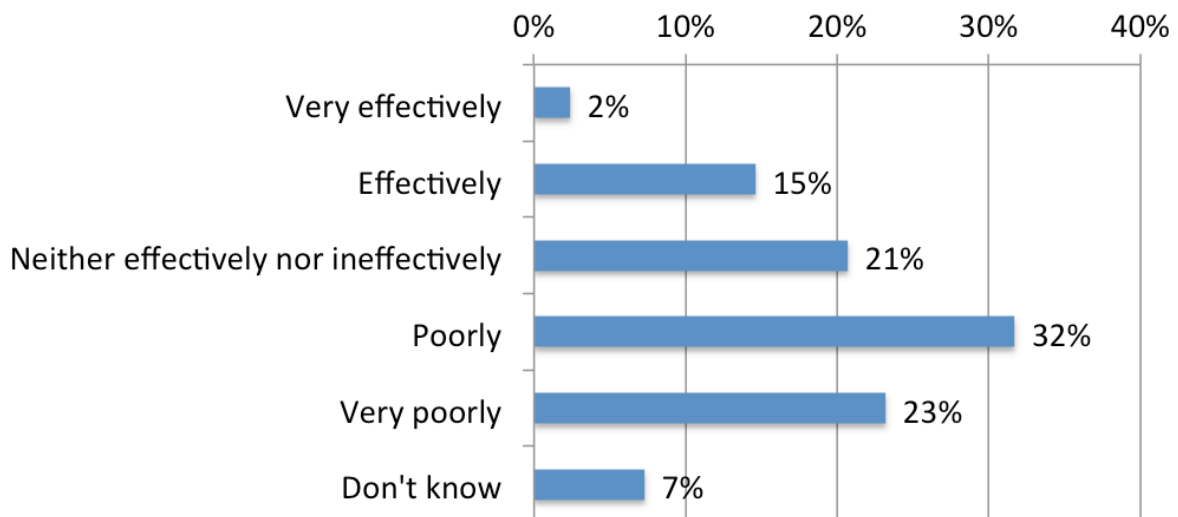


Figure 9 - Ability to handle Unstructured Data

Clearly some 55% are unimpressed with the ability of their data warehouse to deal with unstructured data. Only 17% indicated that they were happy with this. This highlights a clear issue with dealing with such data as that resulting from websites and sensors.

ADOPTION OF BIG DATA SOLUTIONS BY ORGANIZATIONS

We were curious to discover whether respondents considered Big Data to be important for their organization [Q8]. The responses are shown as Figure 10.

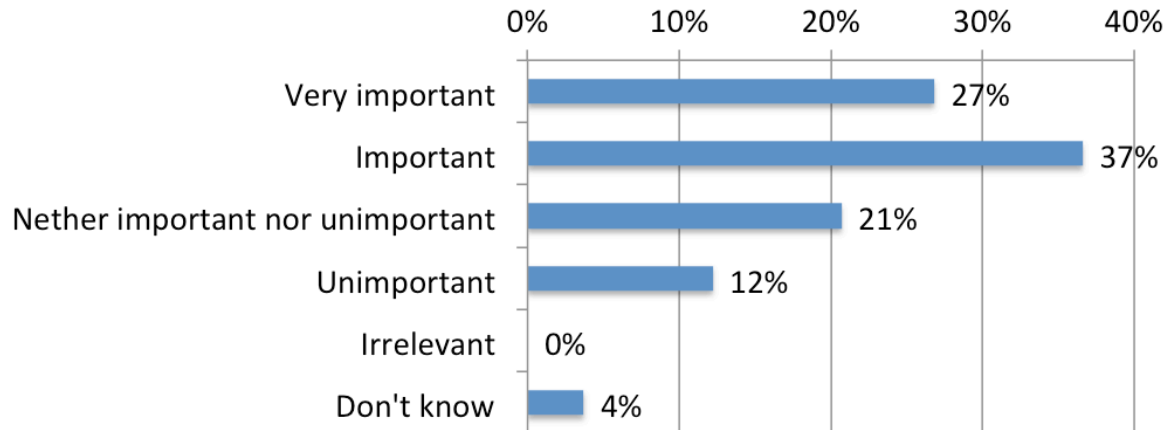


Figure 10 - How important is Big Data for your Organization?

So 64% of respondents consider Big Data to be important for their organization. We conclude that either this is a real issue or organizations are being strongly influenced by the current media hype.

To probe further into the importance and adoption of Big Data we asked respondents to share with us whether they have at least one live Big Data initiative within their organization [Q9]. The feedback is shown in Figure 11.

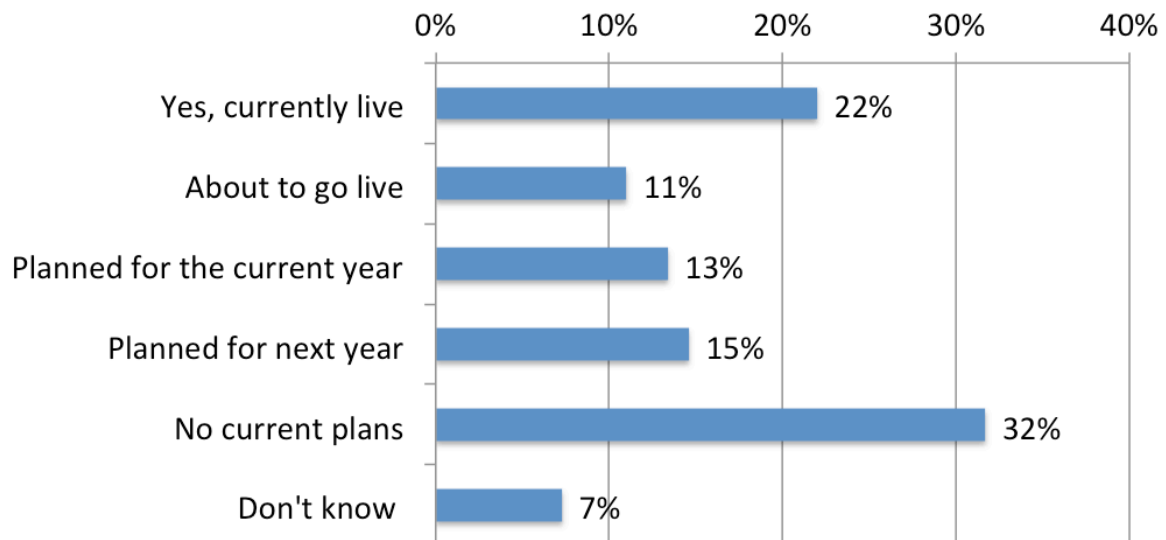


Figure 11 - Have you an active Big Data initiative?

Encouragingly, around one fifth (22%) told us that they already have a Big Data initiative live while a further 11% are about to go live. This is very significant for a technology that is still relatively immature and suggests that many organizations are testing the temperature of the water. What is somewhat baffling, given this significant number of live implementations, is that there appears to be a dearth of case studies available in the public domain. "Where are they?" one might ask?

How big are these implementations in fact? Do they reflect organizations taking some initial steps to explore the potential business value in these approaches? We asked respondents about the size of their implementations [Q10]. Their responses are summarized in Figure 12.

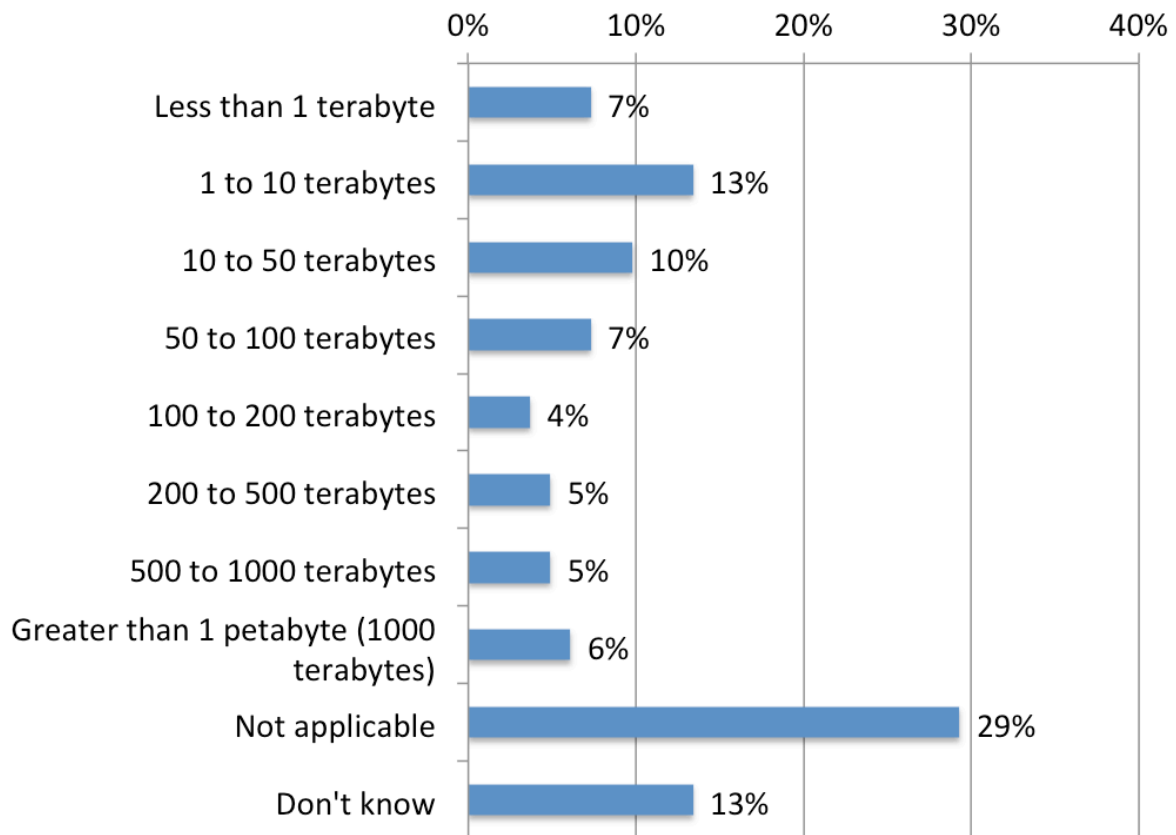


Figure 12 - Size of Big Data Implementations

Around 37% of implementations are less than 100 terabytes with 30% being less than 50 terabytes. These represent really quite small implementations, which prompts us to suggest that organizations are looking more to test out the temperature of the water (and the technology) possibly with a view to further extension. Very few (6%) have implementations in the petabyte range which is where one might expect machine-generated data from sensors or web logs to be placed.

So are these Big Data initiatives any better at coping with unstructured data than conventional data warehouses? We asked respondents to tell us how well they consider their Big Data initiative can cope with unstructured data [Q11]. The results are given in Figure 13.

Amongst those who have implemented Big Data solutions there is a clear view that Big Data technologies are much more effective at helping them analyse their unstructured data. 43% considered these technologies to be at least effective compared with a mere 8% who believed that they were poor at handling unstructured data.

Clearly then the experience is supporting the view frequently expressed in the media that unstructured data is best handled by Big Data solutions such as Hadoop.

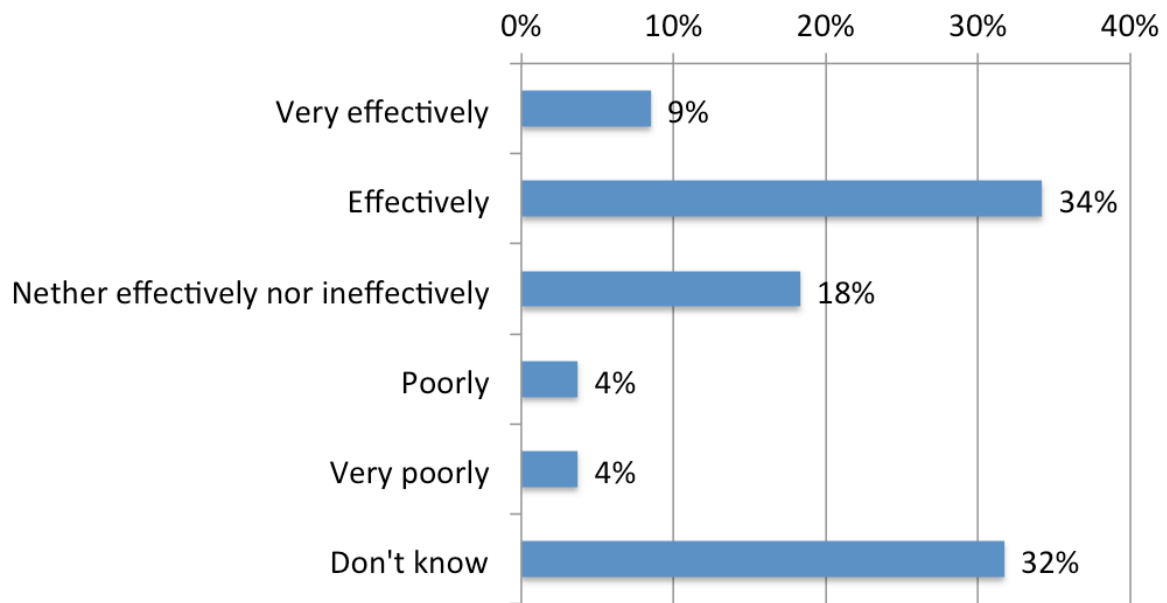


Figure 13 - Ability to handle Unstructured Data

For those who have currently implemented Big Data initiatives, how happy are they with their implementation [Q12]? Is this satisfaction with the ability to process unstructured data reflected in the overall satisfaction with the system? The feedback from the respondents is given in Figure 14.

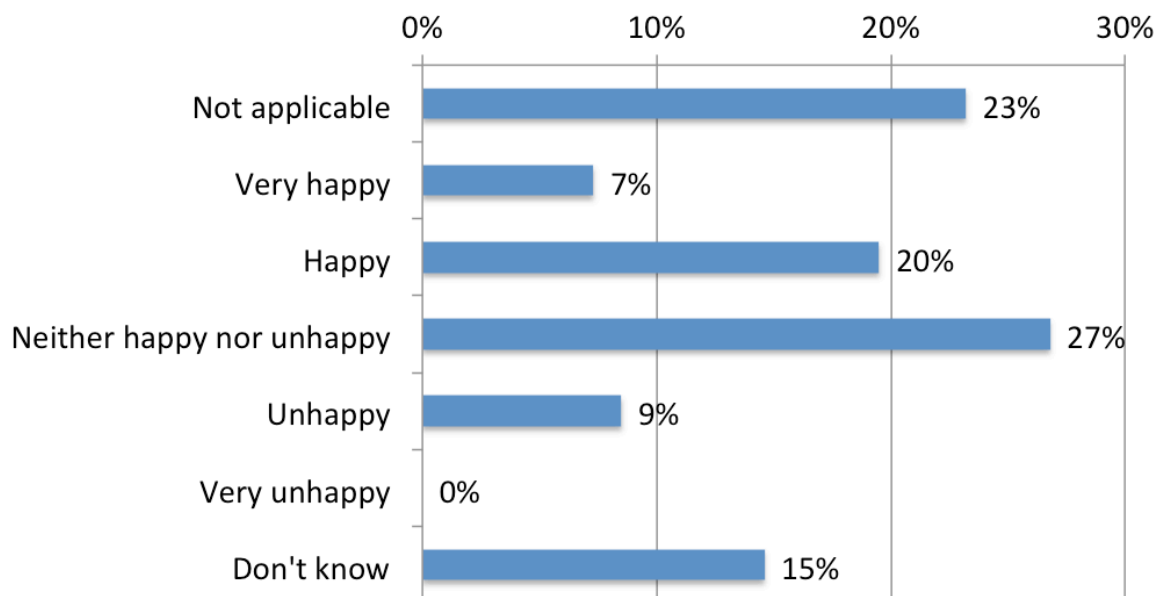


Figure 14 - How happy are you with your Big Data implementation?

The outcome is really quite positive with 27% at least happy with the initiative and just 9% unhappy, a 3:1 ratio. While this is most encouraging it still poses the question why are there so few good case studies available? Are the vendors still busy writing them up or are clients not as yet willing to come out in public with these preliminary results?

THE DEMISE OF THE DATA WAREHOUSE?

So based upon the feedback from the foregoing sections, should we conclude that the demise of the conventional data warehouse is inevitable, close or unlikely?

We asked the organizations to share their view as to whether they considered that Big Data solutions such as Hadoop would eventually take over the role of data warehousing [Q13]. The full results are set out in Figure 15.

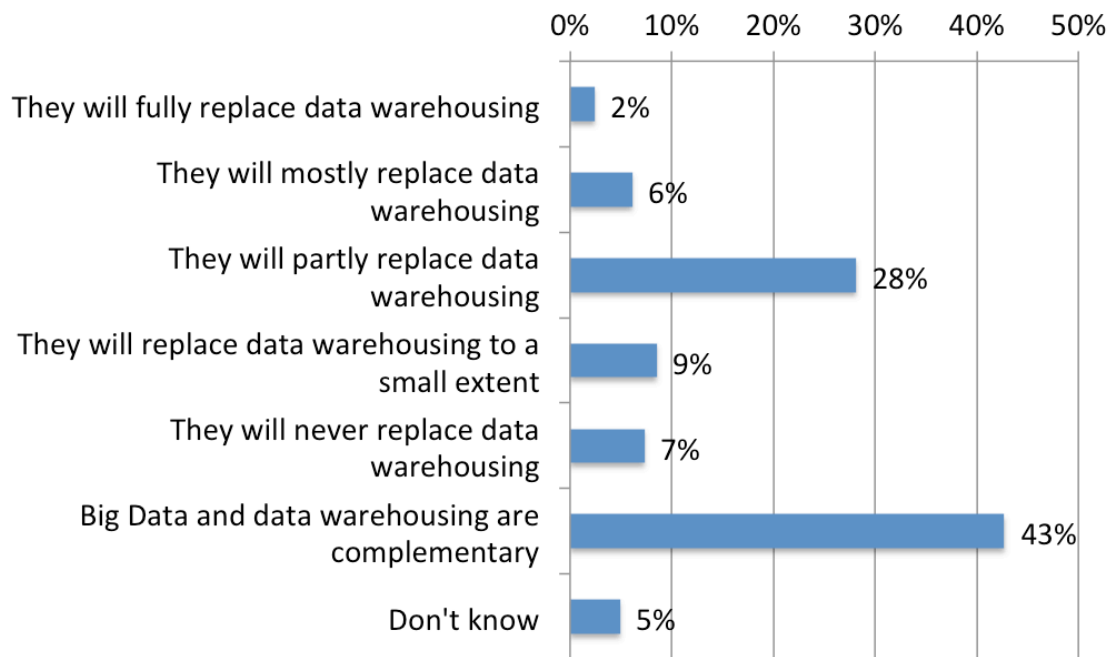


Figure 15 - Will Big Data solutions replace the data warehouse?

Despite the very positive views expressed above concerning the ability of Big Data approaches to effectively handle unstructured data, their overwhelming view is that the technologies are and will continue to be complementary (43%). Some 7% appear to be in denial about Hadoop and the like while only 2% expressed the view that they will replace data warehousing.

Indeed, the roles may be more aligned with the need to use conventional data warehousing to ensure consistency of the underlying data for subsequent use by Big Data solutions. In this context the following quotes [see Q14] seem to highlight some of the key issues:

- *We have enough trouble integrating all the data sources we have now.*
- *Apache Hadoop open source technologies still feel immature compared to conventional data warehousing technologies.*
- *I suspect that "Big Data" is a way to THINK that you are obtaining good data while avoiding the hard work of understanding and designing data models.*

We asked [Q14] respondents to share with us any additional views they had relating to data warehousing and Big Data solutions. The responses included:

- Big Data is for niche analytical / predictive problems.
- Big Data is not only Hadoop, noSQL databases, massive in memory solutions, real time data.
- Big data technologies are complementary to existing data warehouses.

- Complementary technologies - Content vs. Context.
- There are only pilots right now.
- Think the need for the traditional data warehouse will still be required.
- This is important but far from being the top priority.
- We do not have enough data to use the phrase "Big Data".
- There is interest in doing a Big Data initiative but resources are too constrained at this point.
- Hadoop should be part of the solution in the future but certainly at the moment it is not the solution.
- We have enough trouble integrating all the data sources we have now - no one even talks about "Big Data" that I know of.
- We think that Big Data solutions fit better on bottom up discovery (without a predefined model) and search based analytics suitable for use cases where accuracy up to 'almost definite' would suffice. We have a few such use cases. On the other hand, we have many use cases where we need definite analytics based on a predefined model using a data warehouse. In our organization (Which is a \$22bn Joint Venture of a \$240bn company), we implement Information As A Service solution with a logical data warehousing architecture.
- Having a separate Big Data platform will allow us to make decisions faster and will complement our data warehouses.
- In our industry, the Big Data element is primarily driven by the availability of greater varieties of external data. This data often requires partnerships to make it run.
- In today's world of business information needs to accessed at fingertips and also it needs to be presented with all the insights - so if data warehousing does not find a quicker way to minimize the time taken to extract, transform and load data then data warehousing will have a slow death. On the other hand it is too premature for us to comment on the ROI what Big Data has to offer. I see that the Big Data related tools which are used for massaging structured data to (will) become more popular.
- We have had several Big Data initiatives proposed, but never funded. It would be great to augment, but not replace our EDW.
- Data warehousing has reached end of life for most small/medium businesses. The effort outweighs the value obtained.
- Companies believe they need Big Data even when they don't. It's a buzzword that they think they should be adopting. Over time some will shift to Big Data for genuine reasons, some because of the buzz, but the warehouse will persist for the majority who don't need Big Data. Data volumes will grow, but hardware/performance will scale so the proportion that need Big Data will remain the same.
- Big Data solutions are ideal for replacing staging areas in traditional data warehouse solutions. Some analytics can be conducted against them but they are also a good source for data to feed a conventional data warehouse. I think ETL tools need to adapt to fit this new model.
- Our approach is to index data across multiple systems and only bring it together for specific purposes. Even then we use a third party for the analysis.
- Big Data is an overused marketing term for a genre of new generation tools now available that are especially effective in handling, managing, and analyzing massively scaled data set and/or datasets that incorporate a sizable proportion of significant unstructured data. These new generation tools, such as Hadoop, should be considered for use in any comprehensive data warehouse plan to which they are germane. Big Data is a means, not a plan which uses means. Data warehousing is a plan that uses many means, such as "Big Data" tools, to accomplish larger objectives.
- Hadoop/Big Data platforms are an extension of the existing data warehouse environment. Both Hadoop and data warehouse DBMS's will make up a logical data warehouse.

- Our organization must clearly understand that a very large amount of data being stored does not itself imply the need for a Big Data solution. It all depends on how much of the data is NEEDED, not how much we happen to have.
- A data warehouse is still an ideal place for non-RT reporting, there is no need to think it "dead."
- Data processing technologies appropriate to the data and analytics required need to be brought together to solve whatever business problems exist (e.g. in insurance, geospatial, risk management and text analytics would help a broker deciding on the risk of flooding; pattern recognition, web log and time series analysis would help marketers see trends in customer use of a website etc.). Big Data is just one of a number of technologies that need to sit alongside data warehousing technologies.
- Big Data architectures are useful for storing large volumes of unstructured data cheaply. However, doing analysis on that data is still difficult. The learnings (e.g. customer score) you get from such systems, though, need to be stored somewhere where users of any BI tool can get to them. The data warehouse is a good place for that. The data warehouse and/or MDM system could also provide the 'golden copy' of data used for Big Data analysis (e.g. master customer or product list).
- Big Data is the new process to receive new data and offload process for the data warehousing. Big Data is a new source of live/active data to help data warehousing store accurate information historically for effective operations and provide accuracies information for analytics, forecasting and responses to the market trends. Thus Big Data and data warehousing is complementary to each other.
- Big data and data warehousing share the same basic goals: to deliver business value through the analysis of data - thus it will come to which one does it better.
- We are in the middle of piloting and over time this will change dramatically. Today we have to agree with clients on the services they require now and in the future as more moves to virtual.
- Strong business case such as data discovery and predictive analytics is needed for having Big Data initiative. Data warehouse is a necessary component in the operational process of an organization. The concept of integrated, structured data supported by a data model that is designed to enable contemporary reporting and analysis is quintessential to automating and optimizing operational processes.
- Big Data may allow our data warehouse to get rid of things it does poorly and allow it to concentrate on things it can do well.
- Still struggling with practical use-case for Big Data. This is so mainly due to limited or lack of experience of the data warehouse team.
- Hadoop and Relational-technologies are complementary. The trick is to use the right underlying implementation to have a sweet spot on cost/performance. Significant portions of Big Data data require standardization and alignment, which are essentially the data warehouse's domain. It then depends on the use case whether a pure Hadoop approach is sufficient or should be complemented.
- The big problem is our clients must accept that their data will stay in "the cloud" and not in our servers. This slows down the process to switch to manage a data warehouse in a cloud. And the implementation of a Hadoop platform in-house will force you to increase the maintenance costs. We think the future is to manage the data warehouse in the cloud.
- As of now they both are complementary, might change once the maturity of the technologies and offerings evolve.
- I suspect that "Big Data" is a way to THINK that you are obtaining good data while avoiding the hard work of understanding and designing data models.
- In the financial area we need to have a high degree of data governance on traditional structured data driven by regulatory requirements. At the same time we need to be able to handle unstructured data to assess risks and enable customer growth through handling of unstructured

data in huge amounts. This is why we need to be able to handle both traditional data warehouse and Big Data approaches so we can develop synergies and grow our business in a profitable way.

- Big Data will become more important for our company. But at the moment there is no political will for a Big Data initiative as well as for a real enterprise data warehouse.
- Apache Hadoop open source technologies still feel immature compared to conventional data warehousing technologies. There is a disappointing lack of agreed reference design patterns for the tiers in a Hadoop architecture.
- There will always be separation of structured data versus non-structured data. Unless Big Data solutions such as Hadoop can come up with a way to create rich data cubes, there will always be a place for traditional data warehouses and they will coexist with Big Data implementations such as Hadoop.
- Technologies: Big Data should complete IT solutions: data warehouse/BI/ECM. Algorithms and applications, Big Data and data warehouse/BI/analytics should work much closer: both IT and business. Big Data should build a bridge between data-manipulation teams (IT) and information-manipulation teams (collaborative, ECM, KM..) teams. It boils down to bringing together structured and unstructured data people. The CDO function is essential, so there should be a similar CIO/CKO focused on information/knowledge, not the CIO cum technologist/manager/fireman/informatics systems person. This is still needed but either as a service provider or as a manager of internal IT resources (information in the Claude Shannon sense).

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

- 88% of organizations have at least one data warehouse. Most (45%) have between 2 and 5, with all the associated issues of maintaining data consistency.
- Only some 23% have got down to just a single data warehouse. This is despite much media focus on the importance of consolidating to a single data warehouse often linked to some form of master data management (MDM) system.
- 34% have less than 10 terabytes of data suggesting that Big Data may well not be for everyone.
- Most organizations (41%) have data warehouse sizes in the range 1 to 50 terabytes, which is relatively modest.
- The costs of support and maintenance (expressed as FTE) varied with a mean of 19 FTE and a median of 7 FTE. This is a significant level of resource and data warehousing vendors should seek ways to help organizations to reduce this. Our experience suggests that much of this resource is devoted to ETL (extract, transform and load) activities.
- Some 27% of organizations express unhappiness with the costs of their data warehouse.
- In general, organizations are, however, happy (53%) with their data warehouse(s).
- However, 55% were unimpressed with the ability of their data warehouse to handle unstructured data (e.g.: web logs, text, sensor data, ..). This is clearly a stumbling block for many organizations and, possibly, a key reason to look to Big Data technologies.

- Two thirds (64%) of organizations consider Big Data to be important for their business. Either this is a real issue or organizations are being strongly influenced by the current media hype.
- Encouragingly, around one fifth (22%) of organizations already have a Big Data initiative live while a further 11% are about to go live. This suggests that organizations are testing the temperature of the water to evaluate whether the implementations live up to the media hype. Given the number of implementations we are surprised that it remains very difficult to obtain good business case studies. This needs to be addressed by the technology vendors.
- Interestingly, around 37% of the current implementations are less than 100 terabytes with 30% being less than 50 terabytes. This suggests that these are relatively small implementations more akin to pilots.
- Amongst those with live Big Data implementations there is a clear view (43%) that Big Data technologies are much better at handling unstructured data.
- 27% are at least happy with their Big Data implementation with just 9% claiming to be unhappy.
- The general view (43%) is that data warehousing and Big Data technologies are currently, and will remain, complementary. Only 2% claimed that Big Data technologies would replace data warehousing.
- So, the data warehouse is not dead, but very much alive.
- Based upon the broader feedback from the survey, we suggest that the technologies need to be closely aligned with conventional data warehousing (together with MDM) ensuring data consistency for subsequent use by Big Data technologies. This view is highlighted by one quote from the survey: *"I suspect that 'Big Data' is a way to THINK that you are obtaining good data while avoiding the hard work of understanding and designing data models."* This is certainly a potential pitfall.

Vendors

- Just 23% of organizations have consolidated to a single data warehouse. Clearly there is an opportunity here for vendors to engage with their customers to help them to transition to a single consolidated warehouse. This also suggests that there is an opening for MDM and data quality vendors to help provide technology to ensure data consistency.
- Most organizations (41%) have relatively small data warehouse implementations in the range 1 to 50 terabytes. This may also open options for vendor support for consolidation of multiple warehouses.
- Some 27% of organizations are not happy with the costs of maintaining and supporting their data warehouse. There is clearly an issue here to be addressed by warehouse vendors.
- 55% are unhappy with the ability of their data warehouse to deal with unstructured data. Vendors should take this concern on board and seek ways to resolve the position, possibly by incorporating Big Data technologies into their offerings.
- Two thirds (64%) of organizations currently consider Big Data to be important for their organization. This represents a significant opportunity for vendors to engage with organizations to help them understand fully the potential of Big Data. We suggest that the key to success here is for vendors to develop case studies aimed at the business users effectively explaining the business benefits which can potentially result from adopting Big Data technologies such as Hadoop. Currently, we perceive a worrying lack of such case studies.
- Around 37% of current implementations of Big Data technologies are less than 100 terabytes. These are relatively small and suggest to us that they are most likely pilot tests. Vendors should engage with organizations to explore whether this is indeed the case and outline plans for extending the size and scope.

- There is a clear view (43%) that Big Data technologies are much better at handling unstructured data. This is a key focus area for vendors.
- 43% expressed the view that data warehousing and Big Data technologies are complementary, and only 8% thought that Big Data would even mostly replace data warehousing. This suggests that vendors of conventional data warehousing technology would do well to form alliances with Big Data technology vendors, or incorporate Big Data technologies themselves, in order to provide a complete solution approach.

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.

Is the Data Warehouse Dead?

"A data warehouse (DW, DWH), or an enterprise data warehouse (EDW), is a system used for reporting and data analysis. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Data warehouses store current and historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons." (Source: Wikipedia)

"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)

Big Data has generated much interest and attention in the media of late. Indeed, several authors have recently raised the question of whether Big Data approaches, such as Hadoop, will pronounce the death sentence on the conventional data warehouse. At The Information Difference we considered it opportune to explore the views of the data warehouse user community.

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) Do you currently have one or more data warehouses?*

- ☐ Yes, currently live
- ☐ About to go live
- ☐ Planned for the current year
- ☐ Planned for next year
- ☐ None
- ☐ Don't know

2) How many data warehouses have you in your organization?*

- ☐ None
- ☐ 1 (one)
- ☐ 2 to 5
- ☐ 6 to 10
- ☐ 11 to 20
- ☐ 20 to 50
- ☐ More than 50

☐ Don't know

3) What is your overall assessment of your current corporate data warehouse(s)?*

- ☐ Excellent
- ☐ Very satisfactory
- ☐ Neither satisfactory nor unsatisfactory
- ☐ Poor
- ☐ Very poor
- ☐ Don't know

4) How happy are you with the running costs of your current data warehouse(s)?*

- ☐ Very happy
- ☐ Happy
- ☐ Neither happy nor unhappy
- ☐ Unhappy – costs are too high
- ☐ Very unhappy – costs are much too high
- ☐ Don't know

5) Please indicate the approximate size of your data warehouse(s)?*

- ☐ 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

6) Please estimate the number of full-time equivalents (FTEs) that are currently engaged in maintaining your data warehouse(s)?*

7) How well do you consider your data warehouse(s) can cope with handling “unstructured” data (e.g.: text, web logs and sensor data)?*

- ☐ Very effectively
- ☐ Effectively
- ☐ Neither effectively nor ineffectively
- ☐ Poorly
- ☐ Very poorly
- ☐ Don't know

8) Is Big Data important to your organization?*

“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)”

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Irrelevant
- ☐ Don't know

9) Do you have at least one Big Data initiative?*

- ☐ Yes, currently live
- ☐ About to go live
- ☐ Planned for the current year
- ☐ Planned for next year
- ☐ No current plans
- ☐ Don't know

10) Please indicate the approximate size/volume of data handled by 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)
- ☐ Not applicable
- ☐ Don't know

11) How well do you consider your Big Data initiative can cope with handling “unstructured” data (e.g.: text, web logs and sensor data)?*

- ☐ Very effectively
- ☐ Effectively
- ☐ Neither effectively nor ineffectively
- ☐ Poorly
- ☐ Very poorly
- ☐ Don't know

12) How happy are you with your current Big Data initiative?*

- ☐ Not applicable
- ☐ Very happy
- ☐ Happy
- ☐ Neither happy nor unhappy
- ☐ Unhappy

- ☐ Very unhappy
- ☐ Don't know

13) Do you believe that Big Data solutions such as Hadoop will eventually take over the role of data warehousing?*

- ☐ They will fully replace data warehousing
- ☐ They will mostly replace data warehousing
- ☐ They will partly replace data warehousing
- ☐ They will replace data warehousing to a small extent
- ☐ They will never replace data warehousing
- ☐ Big Data and data warehousing are complementary
- ☐ Don't know

14) Please enter below any additional views/comments that you may have in regard to the respective roles of Big Data and data warehousing in your organization.

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

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

16) 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
- ☐ Retail

- ☐ Telecommunications Services
- ☐ Transportation Services
- ☐ Other (Please specify): _____

17) 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

18) Are you willing to take part in a brief, confidential discussion on this topic with an Information Difference analyst?*

- ☐ Yes
- ☐ No

19) Please provide your brief contact details below:*

First Name: _____

Last Name: _____

Company Name: _____

Email Address: _____

20) Please select your country from the drop down list:*