Big Data Trends in 2013

Can You Handle Your Unstructured Data?

February 2013

Nathaniel Rowe
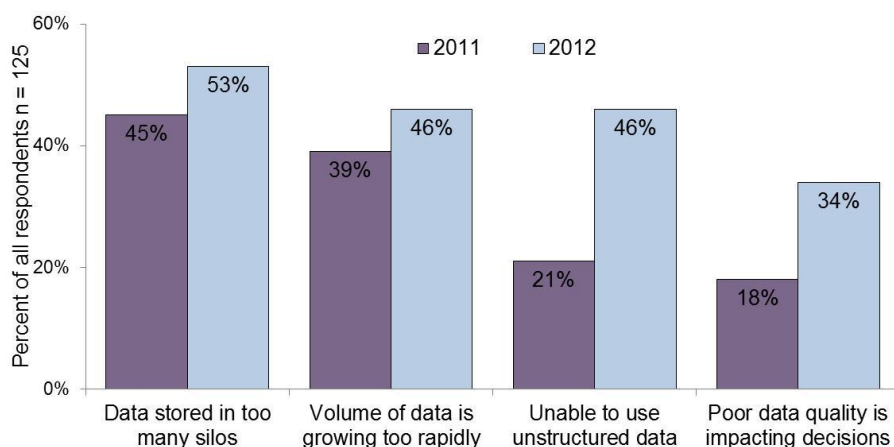
Analyst Insight

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

"Big Data" was one of the hottest technology marketing terms in 2012, resulting in a number of vendors expanding their ability to handle large volumes of data. As the new year begins, Aberdeen looks back at responses from 261 organizations to our January 2012 study and compares them to the latest January 2013 study, comprised of responses from 125 organizations. Big Data challenges continue to rapidly evolve, and as such this report will identify the new pressures and strategies implemented by organizations today. Technology solutions will also be examined in depth, identifying the hot areas for Big Data investment that companies can use to drive business value in the next calendar year.

Rising Concerns Over Big Data

Big Data may be a popular marketing buzzword; however some very real business concerns lay beneath all the hype. In 2011, organizations indicated that their top business pressures included the rising volume of data and the number of silos in which this information was stored. In 2012, anxiety about these problems only escalated, with 18% more organizations reporting that volume and data silos topped their list of data concerns (Figure 1). However, some new concerns, such as data quality, skyrocketed onto the scene in 2012. Over twice the number of organizations reported that their business decisions were negatively impacted by data quality.

Figure 1: Business Pressures Driving Data Investment

Source: Aberdeen Group, January 2013

> **Analyst Insight**
>
> Aberdeen's Insights provide the analyst's perspective on the research as drawn from an aggregated view of research surveys, interviews, and data analysis

> **Definitions**
>
> √ **Big Data** refers to the problems of capturing, storing, managing, and analyzing massive amounts of various types of data. Most commonly this refers to terabytes or petabytes of data, stored in multiple formats, from different internal and external sources, with strict demands for speed and complexity of analysis.

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 2

With data volumes exploding, and on average only three quarters (77%) of business records reported as accurate and reliable, it has become more difficult to sift through the glut of data for nuggets of actionable intelligence. Unstructured data has likewise captured a lot of attention recently. In past years, there was very little that could be done with information residing outside of a traditional relational database. However, with new technology developments around storing, managing, and analyzing different data formats, more companies are looking to expand their data architecture to accommodate these new opportunities.

The following analysis will look at these pressures in greater detail, and highlight the evolution of the modern data infrastructure in relation to the emergence of Big Data.

Deep Dive: More Complex Data Environments

It has been no secret that business data is rapidly expanding, and Aberdeen has followed this trend for several years. The average annual growth rate for data in 2012 is twice the rate reported three years earlier in 2009 (Figure 2). At 56% growth every year, this means that all storage requirements — and associated costs — double every19 months. The data expansion shows no sign of stopping, either. The Best-in-Class companies (see sidebar) in the Aberdeen report on _Data Management for BI_ (January 2013) reported over 100% growth every year, indicating that for companies fully embracing a Big Data philosophy, managing data storage is even more of a pressing issue.

Figure 2: The Data Explosion, 2009 to 2012
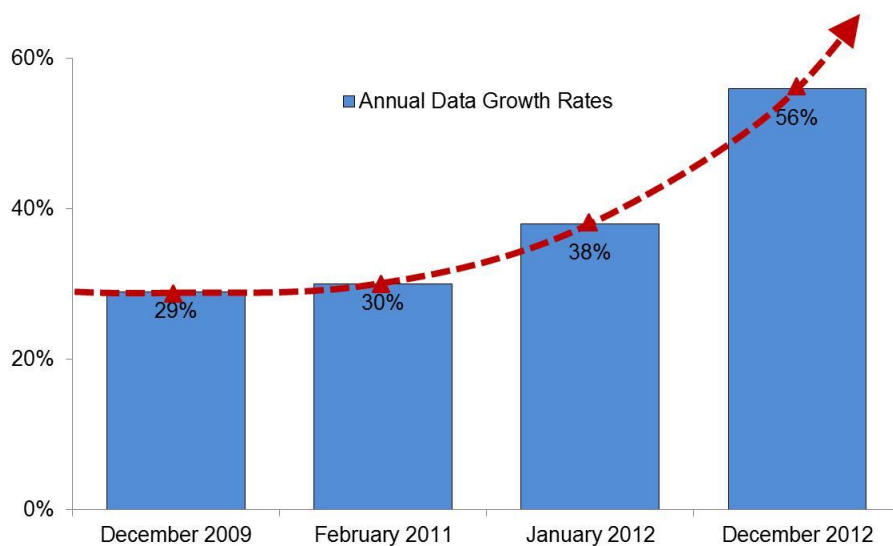
Source: Aberdeen Group, December 2009 – December 2012

However, it isn't just the amount of data that is growing. The diversity of information sources is expanding both internally and externally. In 2011,

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 3

Aberdeen Group
A Harte-Hanks Company

companies averaged 26 unique data sources feeding their applications and analytic systems, but this increased 19% in 2012 to 31 sources. The number of internal sources showed only a slight increase, indicating modest efforts to integrate new databases or software applications. The real growth, however, came in external data. Internet-generated content, such as social media, Twitter feeds, and blogs, are especially valuable, serving as windows into how customers react to products, marketing promotions, and brands. Furthermore, advances in marketing and sales automation have left a lot of companies ravenous for more data in order to optimize their efforts, a niche filled by third-party data aggregators (e.g. economic trends, demographic profiling, customer spend) and content providers. All this information needs to be stored, managed, and controlled, which contributes to the rising cost and complexity of data environments.

A substantial amount of this data is unstructured — 48% of all business data, according to a February 2012 study on *Document Management* — which adds additional headaches for IT. One of the most interesting technologies developed in recent years is called Hadoop, a flexible storage framework that can handle both structured and unstructured files, and distribute the processing and analysis of this information across a networked cluster of servers. Hadoop has become a widely recognized term and essential component of enterprise-level software solutions among the Big Data community, however Aberdeen's research indicates that it has not yet reached the ears of everyone (Figure 3).

Figure 3: Hadoop? Never Heard of Him.
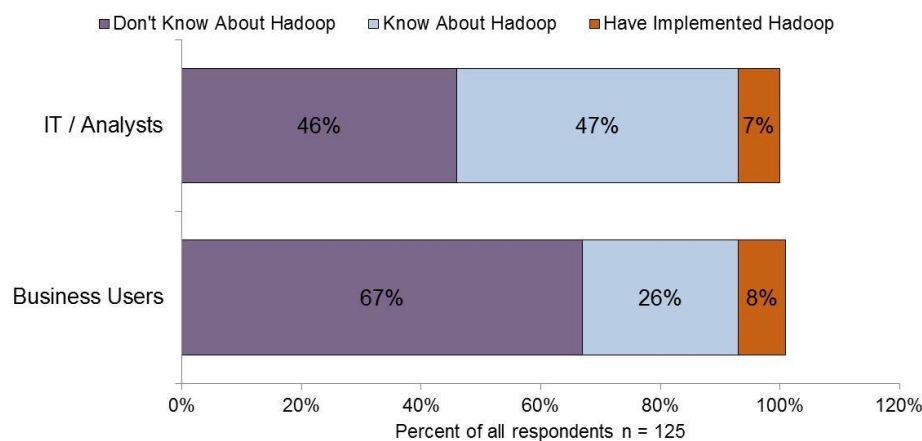
Source: Aberdeen Group, January 2013

The number of respondents in the January 2013 study on *Data Management for BI* who had heard of Hadoop was virtually unchanged since the previous year. This indicates that while Hadoop is very popular in a certain niche of the data management community, it has stayed just in that niche. As Figure 3

Fast Facts

√ **80%** of Best-in-Class companies from Aberdeen's January 2013 report on *Data Management for BI* indicated that they use a significant amount of **unstructured data,** compared with **46%** of all other companies.

√ **Unstructured data** refers to data stored in files, documents, presentations, spreadsheets, web pages, email messages, instant messages, images, audio files, video files, etc. While each of these formats do indeed have "structure," conventional use of the term unstructured data is intended to distinguish from data stored in *structured* formats (e.g. in databases).

"The biggest challenge we have is finding a resource that can help manage unstructured data."

~ Marketing Manager

Small Transportation / Logistics Company

North America

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 4

shows, IT users are much more likely to know about Hadoop and its capabilities than business users, further cementing its status as a hardcore technical term. However, while the specific name of "Hadoop" might not have saturated the market just yet, the need for a management system for unstructured data is well known. For organizations with over five terabytes of data (a standard Aberdeen benchmark for Big Data), only a quarter (24%) indicated that they could manage and use most of their unstructured data, while 21% reported being able to use at least some of it. Among the rest, those that were currently unable to manage their unstructured data, a staggering 50% indicated that this was an area their company was focusing on improving, while only 6% said that unstructured data wasn't a priority at all.

In the end, all the complexity from the growth in data volume, data sources, and data formats impacts not just IT, but day-to-day business operations as well. Finding reliable information and delivering it quickly to managers and decision-makers has become a daunting task, and one that many companies struggle with. Only 27% of the Laggard organizations in *Data Management for BI* (January 2013) reported being satisfied with the accuracy of their business decisions, which helps explain why so many companies listed data quality as a top pressure in 2013. However, the Best-in-Class companies in the same study, the ones that had addressed their issues of data quality, data volume, data silos, and unstructured data, reported a satisfaction rate of 87% with their business decisions. The difference between the two is significant, but thankfully there are a number of solutions available to help bridge this gap.

Hot Technologies for 2013

In order to address their business pressures around Big Data, Aberdeen's research indicates there are several key technologies — in addition to Hadoop — that companies strongly consider investing in (Figure 4).

Given the cost and specific employee skill set required to implement, most of these solutions still have low overall adoption rates, but the number of organizations reporting that they plan to invest in the next 24 months gives a good indication of overall market interest. Three of these tools had enough reported interest that the adoption rates are projected to double or even triple over the next two years. **Real-time integration tools** like Enterprise Service Bus (ESB) are particularly suited for dealing with multiple data silos, and for pulling data from live systems into a data warehouse, without disruption, for fast access and analysis. **In-memory computing** stores data directly in the random access memory (RAM) of a server, thus greatly increasing processing speed and eliminating bottlenecks such as slow storage arrays and clogged network bandwidth.

"In the data management world, it is garbage in… garbage out. You must start with quality data."

~ HR Manager

Mid-sized Marketing Company
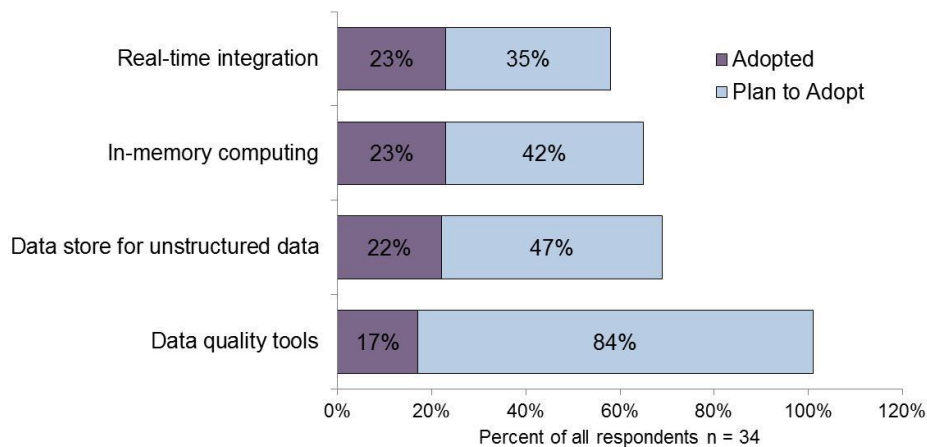
North America

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 5

Aberdeen *Group*
A Harte-Hanks Company

Figure 4: Solutions to Process Data at Speed and Scale

Source: Aberdeen Group, January 2013

Hadoop is the **data store for unstructured data** that has gotten the most attention recently, but there is a growing movement around other not-only-structured-query-language (NoSQL) databases that can handle unstructured data. Aberdeen's March 2012 report, *The Little Elephant in the Big Data World: Hadoop 1.0 Goes Live* delves into greater detail on this technology and its potential business benefits. Finally, **data quality tools** take their place as the solution with the most interest this year, with 84% of companies reporting plans to invest in them. This would more than quintuple the current adoption rate and really drives home the message that it isn't enough to just collect lots of data for a Big Data initiative — this data also needs to be clean, accurate, and reliable. Aberdeen's report *The Big Data Imperative: Why Information Governance Must be Addressed Now* (December 2012) focuses on this essential aspect of data management and highlights some of the hidden costs of poor data quality.

Moving on from the basics of managing data at large scale, Aberdeen's research also shows the technologies being used to take this data and drive real business value (Figure 5). **Social media monitoring tools** have seen impressive levels of adoption in recent years and continue to attract attention. These solutions often collect a massive amount of information, such as all mentions of a product, brand, or company across multiple social media channels, and can correlate those mentions with positive or negative sentiment or specific events. Not only does this allow organizations to better understand which of their products are performing well, but it allows them to quickly identify and respond in the event that negative feedback starts to go viral. **Mobile Business Intelligence (BI) tools** take another popular consumer trend — smartphones — and provide a more internal-oriented business focus. These mobile applications can allow data to be accessed from anywhere, often with data visualization, read / write capabilities, and even rudimentary data exploration. This puts a powerful tool in the hands of executives on the go, field personnel, and remote

"When it comes to data management and Big Data programs, just do it. It may be an expensive change process, but so is lost opportunities and lost business due to poor data quality."

~ CEO

Small IT Services Company

EMEA

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 6

workers, ensuring business data can provide valuable insight regardless of proximity to the home office. While only a handful of companies have such tools in place at the moment, if all companies that plan to invest carry out those initiatives, this adoption rate will multiple by a factor of six in just two years.

Figure 5: Technologies to Drive Value from Big Data
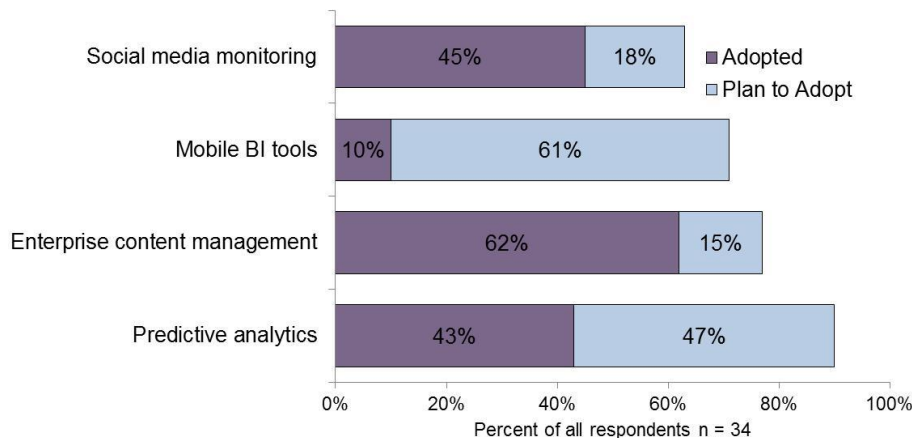
Source: Aberdeen Group, January 2013

Enterprise content management (ECM) provides a platform for an organization to manage and distribute all their important information assets, including documents and rich media files (e.g. audio, images, video). The internal business use cases often revolve around coordinating digital versions of documents, like loan applications or insurance claims. These platforms can also play a crucial role in fostering collaboration among business units, business partners, or even directly with customers through online web portals. Finally, **predictive analytics** takes the traditional BI practice of analyzing the past through historical transactions and moves into the future. Given enough data, predictive analytics can identify patterns and trends in business activity and the larger economy, and with these patterns, identify potential problems before they occur or project the likelihood of lucrative business opportunities. Both these solutions have been adopted by a large percentage of organizations already, but the continued interest shows that companies without these solutions recognize they are at a disadvantage compared to their peers, and are actively seeking to remediate this problem.

Key Takeaways

Whether an organization is on the forefront of the Big Data evolution, or is just now beginning to struggle with the data explosion, Aberdeen's research indicates that there a few key data management issues to address. The

"The Board of Directors have to be persuaded to treat data as an asset, from acquisition, safeguarding, ownership, through disposal. When there is a disconnect between top management and the users of data (particularly unstructured), the data strategy becomes unfocused…"

~ IT Director

Mid-sized IT Services Company

Asia/Pacific

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 7

following have long been popular drivers of investment, but for 2013, they have risen to new prominence:

- **Manage the growing volume of data**. The average organization reported that their data is expanding at 56% year-over-year, meaning all the costs, headcount, and other resources devoted to data storage double every 19 months. Before these support costs get out of control, organizations should take care to develop a sustainable, scalable data strategy.

- **Strategically integrate new data sources**. The average number of unique data sources grew 19% in the last year, and organizations are increasingly concerned that their data is spread out in too many locations. Connecting and centralizing this data is a key step in maintaining a cohesive picture of business operations — as is tactically adding new, external sources of valuable data.

- **Build a strong data infrastructure**. Managing a large volume of data in a variety of different formats requires more tools than just relational databases. Managing unstructured data and ensuring data quality at scale were two critical areas where concern had doubled in the past year. Unstructured data stores and data cleansing tools have therefore become highly in-demand technologies.

- **Find the business value of Big Data.** Best-in-Class organizations in the January 2013 report on *Data Management for BI* were analyzing more data, in more formats, from more sources, at a faster pace than their peers. As a result, they reported a 31% year-over-year improvement in the accuracy of their business decisions. They also reduced the time, manpower, and resources for data-centric tasks like processing invoices or managing inventory by 16% in 2012. Identify which business processes in your company could benefit from better data management, or how a Big Data initiative could add value to your analytics and BI programs. Discovering customer trends through social media monitoring, or streamlining internal processes through ECM tools are two solutions that companies plan to use in 2013 to drive value, but they are far from the only available options.

Big Data Trends in 2013: Can You Handle Your Unstructured Data?

Page 8

Aberdeen *Group*
A Harte-Hanks Company

For more information on this or other research topics, please visit
www.aberdeen.com

Related Research	
Data Management for BI: Getting Accurate Decisions from Big Data; January 2013	*Enabling Access to Big Data with Data Integration*; April 2012
The Big Data Imperative: Why Information Governance Must be Addressed Now; December 2012	*High Performance Organizations Empower Employees with Real-Time Mobile Analytics*; April 2012
Big Data for Small Budgets; December 2012	*Mobile BI 2012: Accelerating Business on the Move*; March 2012
Go Big or Go Home? Maximizing the Value of Analytics and Big Data; September 2012	*The Little Elephant in the Big Data World: Hadoop 1.0 Goes Live*; March 2012
The State of Big Data: Video Benchmark; July 2012	*Divide and Conquer: Using Predictive Analytics to Segment, Target and Optimize Marketing*; February 2012
Agile or Fragile? Your Analytics, Your Choice; July 2012	*Operational Intelligence - Part 1: Driving Performance with Tactical Visibility*; February 2012
Beyond Agile Analytics: Is Agile Data Integration Next; June 2012	*In-memory Computing: Lifting the Burden of Big Data*; January 2012
Managing the TCO of BI: The Path to ROI is Paved with Adoption; May 2012	*The Role of Big Data Analytics in HR: Speed, Satisfaction and Scale*; January 2012
Author: Nathaniel Rowe, Research Analyst, Enterprise Data Management (nathaniel.rowe@aberdeen.com)	