Aberdeen Group
A Harte-Hanks Company

Data Management for BI

Getting Accurate Decisions from Big Data

January 2013

Nathaniel Rowe

Research Brief

Aberdeen Group
A Harte-Hanks Company

January 2013

Data Management for BI: Getting Accurate Decisions from Big Data

Aberdeen's latest research into data management (December 2012) has uncovered the Best-in-Class practices for delivering fast, reliable information to decision-makers. In a survey of 125 organizations world-wide, these top performers were defined by their success at managing big data, such as the ability to collect data from more sources, analyze larger volumes of data in less time, and maintain higher levels of data quality. Through the technological tools, business capabilities, and skilled employees that support these business intelligence (BI) programs, the Best-in-Class companies were able to realize incredible performance boosts in decision accuracy, quality of data analysis, and overall business process efficiency. This report will cover the range of business capabilities and technology solutions being used by these top performing organizations to achieve their analytic success.

Faster Insight, Better Decisions

Understanding how your organization operates and how your customers behave is a barebones necessity for running a company. When you can access company performance data faster, more nimbly react to business events, and be more accurate in your decision making than your competitors, then your company can begin to distinguish itself in the marketplace. According to the 125 organizations surveyed in Aberdeen's December 2012 Data Management research, 56% of the Best-in-Class reported they were using faster, more complex analytics to gain a competitive advantage over their peers. However, this type of advanced, high-speed analytics requires both high-quality data and solid data-management techniques to be successful.

Transforming raw data into timely insight is at the core of a good BI strategy, and doing it quickly even with high volumes of data is the mark of a good big data initiative. Best-in-Class companies from this study were identified based on their ability to bring new data sources into their analytical infrastructure quickly, make more of these data sources accessible, provide more accurate business data, and deliver insight within the time required to take action. Aberdeen determined Best-in-Class performance based on the following four key performance indicators:

- **Data "On-Boarding" Efficiency:** Measured as an average number of days required to integrate new data sources into the organization's information infrastructure.
- **Data Accessibility:** Measured as an average year-over-year increase in the amount of business data that is discoverable or searchable.

Research Brief

Aberdeen's Research Briefs provide a detailed exploration of a key finding from a primary research study, including key performance indicators, Best-in-Class insight, and vendor insight.

Rapid Growth of Data

Business data is increasing at a rapid pace, and Aberdeen has been tracking the trend over the past several years:

- √ **29% growth** year over year was reported in December 2009
- √ **30% growth** was reported in February 2011
- √ **38% growth** was reported in January 2012
- √ **55% growth** was reported in the most recent survey in December 2012

- **Data Reliability:** Measured as an average percentage of business data considered to be reliable and accurate.
- **On-Time Information Delivery:** Measured as an average percentage of actionable business information delivered on-time, or within the required "decision window."

Table 1 shows the average performance of the three maturity classes in these four key metrics.

Table 1: Top Performers Earn Best-in-Class Status

Definition of Maturity Class	Mean Class Performance
Best-in-Class: Top 20% of aggregate performance scorers	▪ **9** days required to integrate new data sources ▪ **35%** year over year *increase* in accessible business data ▪ **93%** of business data is considered to be accurate ▪ **91%** of key business information delivered on-time
Industry Average: Middle 50% of aggregate performance scorers	▪ **58** days required to integrate new data sources ▪ **13%** year over year *increase* in accessible business data ▪ **77%** of business data is considered to be accurate ▪ **76%** of key business information delivered on-time
Laggard: Bottom 30% of aggregate performance scorers	▪ **137** days required to integrate new data sources ▪ **10%** year over year *decrease* in accessible data ▪ **57%** of business data is considered to be accurate ▪ **47%** of key business information delivered on-time

Source: Aberdeen Group, December 2012

Business Drivers for Data Management

Aberdeen's research uncovered several business pressures that prompted organizations to invest in their data management and BI initiatives. However, these drivers showed a significant split between the maturity classes (Figure 1). Given that data management is an iterative process that builds upon previous successes — a walk, crawl, run journey — the challenges faced by companies just starting a data management program will be very different from those faced by an advanced organization with a mature data management initiative capable of handling high speed analytics on large-scale data. The Best-in-Class pressures are easy to understand when presented in the context of their complicated data environments. On average, they collect data from more sources, internally and externally, than other organizations. This is especially true of external, high-volume data sources like clickstream data or social media monitoring.

Fast Facts: Data Accuracy

Best-in-Class companies reported:

√ **94%** data accuracy was their organizational goal.

√ **1%** improvement is necessary for them to meet this goal

Industry Average reported:

√ **91%** was their data accuracy goal

√ **18%** improvement is needed to achieve it

Laggards reported:

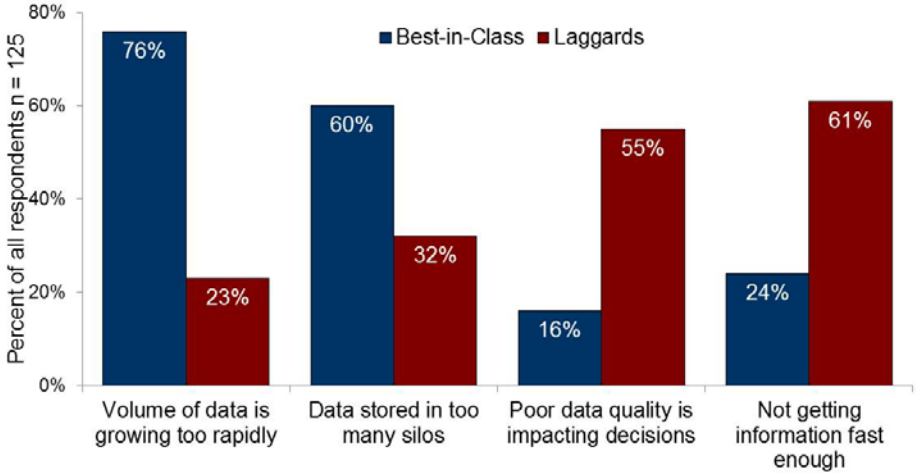
√ **80%** was the low bar they set for data accuracy

√ **40%** improvement in their current performance is still required to meet this benchmark

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.

Figure 1: Business Pressures Vary by Maturity Class



Source: Aberdeen Group, December 2012

"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

The Best-in-Class reported accessing over 23 unique external data sources, which is almost twice the rate of Laggards, who reported only 12. With more data sources comes more data to store, which explains the 76% of the Best-in-Class that listed the growing volume of data as a top pressure. Even the overall rate of 56% data growth year-over-year is enough to give most companies pause, but the Best-in-Class reported an average of 100% annual data growth, meaning their data storage requirements were doubling every year.

Laggard organizations, on the other hand, were more concerned with basic capabilities such as delivering information in a timely fashion and ensuring the data is reliable enough to use in business decisions. The Best-in-Class, however, are further along in their data management journey, which is why they were up to 3.4-times less likely to report these problems as a top pressure.

Analyzing Data at Speed and Scale

What makes the Best-in-Class' success rate at delivering accurate information in a timely fashion even more impressive is the scale at which they are doing it. As Table 2 shows below, these top performers were storing more information overall, were able to analyze a larger percentage of this information, and accomplish this at faster speeds than other organizations. On average, the Best-in-Class were storing over 100 more terabytes of data than Laggards, and analyzing twice the percent of this information, despite being smaller companies on average (see sidebar). This equates to 3.4-times more data being regularly analyzed, giving these top performers a more comprehensive view of their operations and performance. Furthermore, 46% of the Best-in-Class reported they require real-time access to this information, which is over 4-times the rate of Industry Average and Laggard organizations combined. Not only do they need this information faster, they are successfully meeting this demand — as

Fast Facts: Company Size

The average size of the different maturity classes was reported as:

- √ **1,500 employees** for the Best-in-Class
- √ **14,000** employees for the Industry Average
- √ **3,600** for the Laggards

Table 1 (page 2) shows, the Best-in-Class deliver information within this window 91% of the time.

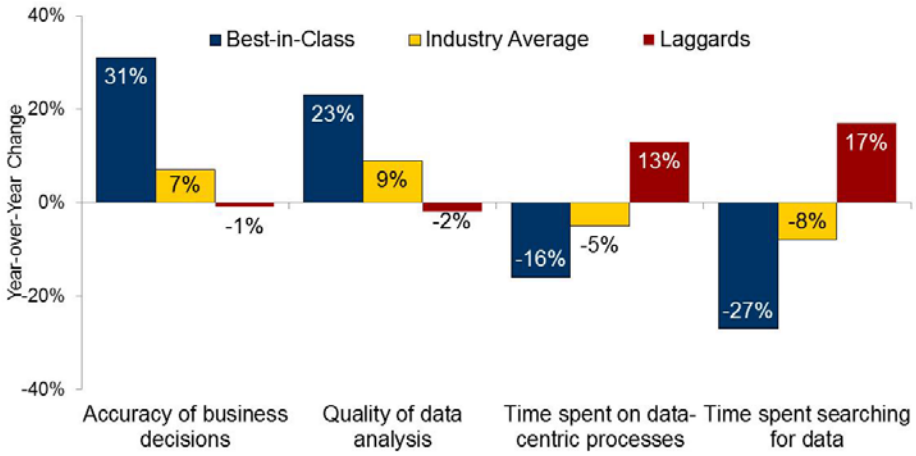
Table 2: Larger Data Volumes Accessed at Greater Speeds

Performance Metrics	Best-in-Class	Industry Average	Laggard
Average amount of business data	240 terabytes	190 terabytes	135 terabytes
Percent of all data available for analysis	37%	22%	19%
Average amount of data analyzed	89 TB	42 TB	26 TB
Percent of respondents needing real-time data access	46%	18%	14%

Source: Aberdeen Group, December 2012

As one might guess, being able to analyze larger volumes of accurate data in a shorter amount of time can contribute to better business performance. The data management techniques practiced by the Best-in-Class directly correlate to their performance in faster business decisions and efficiency (Figure 2).

Figure 2: Annual Change in Decision-Making and Efficiency



Positive result indicates increase; negative result indicates decrease; n = 125

Source: Aberdeen Group, December 2012

In the last fiscal year, the top performers reported a 23% improvement in the quality of their data analysis. Coupled with their already exemplary ability to deliver this insight in a timely fashion, they likewise reported a 31% improvement in the accuracy of their business decisions. On a high level,

Fast Facts: Data Ownership

When identifying what business units "owned" data management initiatives, companies reported:

- √ **44%** were owned by IT
- √ **20%** were owned by a specific senior executive
- √ **12%** were owned by a cross-functional team of IT and line-of-business managers
- √ **10%** were owned by a line-of-business unit

this means that when a disruptive business event occurs, the company can understand exactly how it will affect business operations and quickly take steps to minimize the impact. For instance, consider the tragic tsunami in Japan from a business angle. The natural disaster significantly altered international shipping operations for months, and resulted in severe material and parts shortages. If a company could quickly understand which of their suppliers were affected, which products would be impacted by missing material, and the availability of secondary suppliers, they could be first in line to sign a contract for high-demand resources and keep their operations running. The fiscal implications of such a situation are staggering.

On a more tactical level, Best-in-Class organizations also reported significant improvements in business efficiency. They spent 27% less time searching for the data they needed, and all their data-centric tasks hummed along at a 16% better clip than last year. Conversely, the Laggard organizations, due to low quality data and poor data management, showed massive performance decreases of 17% and 13% respectively. With business data growing so quickly that storage requirements double every 19 months, these companies are becoming overwhelmed by the complexity of storing, accessing, and analyzing this massive volume of information. As we will see in the following section, the Best-in-Class have adopted powerful, scalable tools to manage this big data. As a result, not only are they showing better baseline performance, but they also rapidly increase the gap between themselves and their competition.

Roadmap to Best-in-Class Performance

Given the wide variety of data sources and formats available to organizations, no single technology, strategy, or set of skills will handle all of a company's data needs. A successful data management program includes crucial elements such as skilled and dedicated employees, adequate resources, management support, an organizational culture that values fact-based research and decision-making, and the technological tools to ensure even large volumes of data are accurate, consist, and easily accessible. Even when all these are in place, organizations must remain vigilant. With new data sources constantly added, new data formats needing to be analyzed, and demand for faster and faster analysis, data management is never finished.

"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."

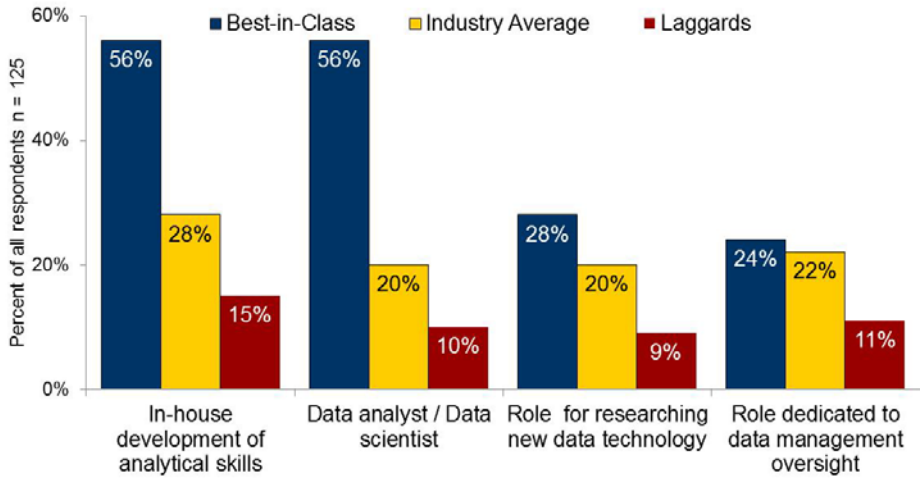
~ CEO

Small IT Services Company

EMEA

One of the first areas organizations should address is having the proper skill set among their employees. On a very basic level, employees in key management positions should understand what data is relevant to the operations they oversee and be trained on the systems used to deliver this information. The next step is to develop more advanced analytic skills for not just digesting data, but exploring and mining data sets to understand problems and develop concrete solutions. As Figure 3 shows, Best-in-Class companies are 3.7-times more likely than Laggards to work on developing these skills among their current employees. Not only does this help with career development and make employees more valuable to the organization, but it puts analytic power in the hands of people who already have a deep understanding of the company and their business operations.

Figure 3: Acquiring an Analytic Skill Set



Source: Aberdeen Group, December 2012

There are certain analytic tasks, however, that require specialized skills in software development and programing languages, or advanced degrees in statistics. Some of the most powerful tools today, like distributed Hadoop clusters and mobile applications, require extensive programming expertise and years of IT experience. When it is impractical to train existing employees, the Best-in-Class are not afraid to reach out and hire these experts. Fifty-six percent (56%) have a dedicated role for data analysts or data scientists — mathematicians, social scientists, or economists devoted to analyzing business data — compared to a mere 10% of Laggard organizations.

Other dedicated roles correlate to Best-in-Class performance. The technologies around data management and Big Data evolve rapidly, and companies that don't want to be caught behind the adoption curve must take proactive steps to research and evaluate new technologies. Finally, having a Chief Information Officer or other management position directly responsible for enterprise data quality is a practice that the Best-in-Class are twice as likely to have in place as Laggards.

"As a data mining specialist, my key takeaway is that a Data Scientist must really understand the business prior to building any Big Data process. To be effective in building analytical solutions that meet the needs of the organization, you really need to have an intuitive grasp of how the operational processes function. The same is true for Big Data processes in the sales / marketing world. The first step must be getting a full assessment of the business itself — whether that's an operational process or understanding the market scope for a given product."

~ Data mining specialist

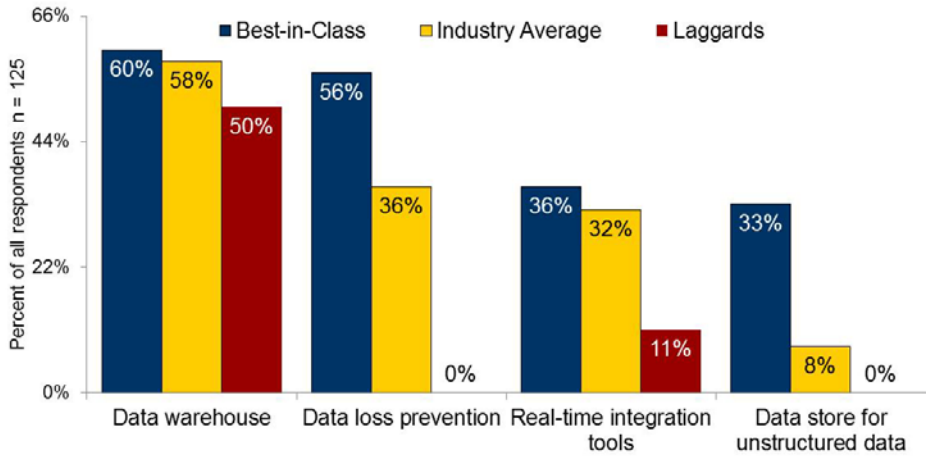
Mid-sized Transportation / Logistics Company

North America

Essential Data Management Tools

In addition to the human element, there are a number of technology solutions that contribute to data management success. The first priority is to address data quality and accuracy, and to consider how these solutions are able to scale up to large data volumes. Given that data is growing at over 56% year-over-year on average, any solution that relies heavily on manual efforts can quickly become impractical and a burden on employee's time. Best-in-Class organizations have turned to **automated tools** to address their data quality; 42% can automatically capture, classify, index, and cleanse data entering their system. None (0%) of the Laggards in this survey reported having these tools.

Figure 4: Storing and Protecting Data



Source: Aberdeen Group, December 2012

On the other hand, **data warehouses**, or separate repositories used to house data exclusively for the purpose of analysis, are tools that are seeing high levels of adoption from all companies, regardless of maturity class. Given the complications inherent in analyzing data that is used live in critical enterprise applications, 50-60% of all companies have established this warehouse to support their BI tools.

The Best-in-Class, however, take the extra step to make sure this data is protected and up-to-date. **Data loss prevention (DLP)** tools can actively monitor business data and detect when sensitive information like credit card or social security numbers are improperly accessed or moved. They can then flag the event, notify an administrator, or block the action entirely. While these tools have high adoption rates among the top performers, the Laggards have ignored this aspect of data security. Furthermore, the Best-in-Class ensure that their data warehouse is filled with the most current version of data possible. They are more than 3-times more likely than Laggards to be able to **integrate** with multiple data sources and move data between them in near **real-time speeds**. Instead of updating a data warehouse once a day or week, the Best-in-Class are more likely to have an almost perfect clone of their live data fueling their BI dashboards.

Finally, unstructured data (see sidebar) is becoming a more important part of business analytics. Unlike structured data that is stored neatly in the rows and columns of a relational database, these text files, emails, videos, images, and social media posts can be tricky to index and store, and even trickier to effectively analyze. There is a growing movement toward NoSQL (not only structured query language) databases that are flexible enough to manage these data formats, and the Best-in-Class are at the forefront of the adoption curve. A third (33%) reported adoption of these or similar technologies to **manage their unstructured data**. For more information

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

~ Marketing Manager

Small Transportation / Logistics Company

North America

Fast Facts

√ **80%** of Best-in-Class companies report that they use a significant amount of **unstructured data**

Compared with:

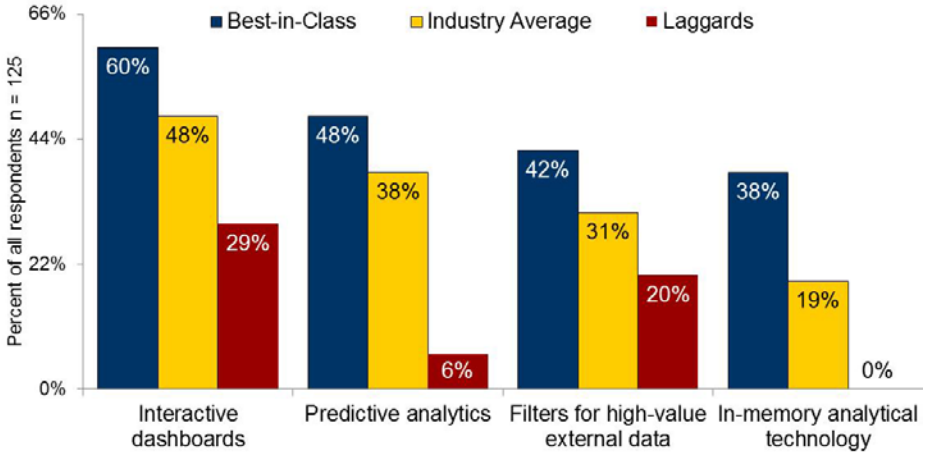
√ **46%** of all other companies

√ **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).

on this topic, see Aberdeen's research into *Apache Hadoop* (March 2012) and *Big Data for Small Budgets* (December 2012).

After integrating data sources and storing data in a warehouse, the next step becomes analyzing this data and delivering it to end-users. The Best-in-Class are more than twice as likely as Laggards to first apply some simple **filters on incoming data**, in order to prioritize the data sources with the most high value information (Figure 5). While not having enough data for analysis can be a problem, so can having too much. When tons of information is stored and available for access, it can easily become overwhelming to end-users, and valuable insights end up concealed under layers of useless statistics.

Figure 5: Analyzing Data and Delivering Insight



Source: Aberdeen Group, December 2012

Likewise, in order to make the end-user experience as easy and engaging as possible, 60% of the Best-in-Class have implemented **interactive dashboards**. Not only are these visual displays easily customizable in how they display summaries of data, they allow employees to explore the data on their own. Without these tools, an employee that sees a report on a store suffering a dramatic loss in sales would have to contact IT to get them to provide a more granular view of the data, or assign another employee to research the situation. With interactive dashboards, end-users could, for example, move from summary information directly down to specific data on individual product sales, shipments, employee turnover, and theft. These tools often provide the first step towards understanding cause and effect.

The Best-in-Class are also more likely to have invested in specific advanced technologies to support their analytic efforts. **In-memory computing** stores data directly in the random access memory (RAM) of a server, instead of having to move it to and from traditional disk storage, allowing for incredibly fast computation. In Aberdeen's January 2012 report *In-Memory Computing: Lifting the Burden of Big Data*, users of this technology

Data Sources to Target

The following data sources were listed as Important or Very Important to Big Data analytic initiatives:

- √ **96%** needed transactional application data
- √ **88%** needed other internal sources of structured data
- √ **85%** needed data on customer purchase history or behavior
- √ **77%** needed external unstructured data from customers or business partners
- √ **68%** needed internal sources of unstructured data
- √ **60%** needed internet-generated data like clickstream and web traffic data

reported 100-times faster data analysis. Lastly, the Best-in-Class not only provided reports on historic information, but also used **predictive analytics** to mine their data for major trends and use these patterns to predict market trends, customer behavior, and product sales.

Organizational Culture and the Value of Data

Perhaps the most difficult aspect of a data management initiative is affecting the organizational culture and how employees view data. For some companies, data is just a cog in the machine, or an obstacle that needs to be overcome for day-to-day tasks. The Best-in-Class, however, see data as an enabler of efficiency, and a valuable asset that provides both better decisions and better company performance. Getting from the first situation to the second is no easy task, requiring time, effort, and eventual buy-in to the philosophy by managers and knowledge workers. Aberdeen's research shows that first and foremost, there is a direct correlation between visible support from management and an increased trust in business data (Table 3).

"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

Table 3: Benefits of a Data-Driven Culture

Category	Best-in-Class	Industry Average	Laggard
High* trust in business data	**96%**	48%	33%
High trust in data systems	**87%**	55%	20%
High adherence to data policies by end-users	82%	43%	**0%**
High support from senior management	74%	40%	**7%**

"High" indicates a rating of 7 or higher on a scale of 1 (low) to 10 (high)

Source: Aberdeen Group, January 2012

One a scale of 1 (low) to 10 (high), almost three quarters (74%) of the Best-in-Class rated the support for data and BI programs from their senior management as a 7 or higher; only 7% of Laggards reported such high levels. Having executives that understand the importance and value of data is critical to getting programs started and technology funded. However, buy-in doesn't end in the board room — it is just as important for end-users to adhere to corporate policies on data creation, security, and management. If no one cares about the quality or integrity of data they routinely accesses, issues of trust will never get resolved. None (0%) of the Laggards ranked their employee's adherence to data policies higher than a 6.

Having this support from all levels of the organization is the foundation of a data-driven culture. With this support comes trust, both in the data systems that are delivering information, and the in the information itself. Ninety-six percent (96%) of the Best-in-Class reported high levels of trust in their data.

With that trust comes the business performance, such as reported earlier in Figure 2: better decisions, quality analysis, less time searching for data, and more efficient processes. And with trust and a proven track record of operational success comes a company willing to invest the time and effort in maintaining and improving their data management programs for years to come.

Summary and Key Takeaways

Aberdeen's research has shown that the top pressures driving organizations to improve their data management are driven by the current big data phenomenon and include expanding data volumes, too many data silos, poor quality data, and slow access to information. For organizations that struggle with these pressures, or are looking to gain a competitive advantage through analytics, Aberdeen recommends the following:

- **Show support for a data-driven culture**. True success comes when both management and the line-of-business employees buy into the idea that data is a valuable asset and should be treated as one. Regardless of the position, becoming involved in data improvement programs and being serious about data quality and data discipline will help set the example for others. Best-in-Class companies were 10.6-times more likely to have strong support from management, and were rewarded by a 31% year-over-year increase in the accuracy of their business decisions.
- **Develop or hire skilled employees**. Analytic skills are critical to understanding data and developing new insights. The Best-in-Class are 3.7-times more likely than Laggards to train up existing employees in these techniques, and over 5-times as likely to hire skilled professionals for advanced analytic tasks.
- **Build a flexible, high-speed data infrastructure**. Feeding data warehouses in real-time, from multiple data sources including unstructured data, are all marks of Best-in-Class performance. Despite the rapid growth of data, these top performers analyzed 3.4-times more data than Laggards, and still delivered it in a timely fashion 91% of the time. Furthermore, an agile infrastructure that integrates new data sources and delivers information at real-time speeds is able to put the right information at the fingertips of decision makers, and allows for faster, more accurate reactions to business events.
- **Provide self-service exploration and advanced analytics.** Putting the tools for data exploration and drill-down capabilities directly into the hands of the end-user not only allows them to quickly answer their own questions, but removes a burden normally placed on IT. Furthermore, embracing tools like predictive analytics, used by 48% of the Best-in-Class, can directly contribute to predicting market trends and getting a competitive advantage.

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

Related Research	
The Big Data Imperative: Why Information Governance Must be Addressed Now; December 2012	*Mobile BI 2012: Accelerating Business on the Move*; March 2012
Big Data for Small Budgets; December 2012	*The Little Elephant in the Big Data World: Hadoop 1.0 Goes Live*; March 2012
Go Big or Go Home? Maximizing the Value of Analytics and Big Data; September 2012	*Divide and Conquer: Using Predictive Analytics to Segment, Target and Optimize Marketing*; February 2012
The State of Big Data: Video Benchmark; July 2012	*Operational Intelligence - Part 1: Driving Performance with Tactical Visibility*; February 2012
Agile or Fragile? Your Analytics, Your Choice; July 2012	*In-memory Computing: Lifting the Burden of Big Data*; January 2012
Beyond Agile Analytics: Is Agile Data Integration Next; June 2012	*The Role of Big Data Analytics in HR: Speed, Satisfaction and Scale*; January 2012
Managing the TCO of BI: The Path to ROI is Paved with Adoption; May 2012	*Data Management for BI: Big Data, Better Insight, Superior Performance*; January 2012
Enabling Access to Big Data with Data Integration; April 2012	*Agile BI: Three Steps to Analytic Heaven*; March 2011
High Performance Organizations Empower Employees with Real-Time Mobile Analytics; April 2012	*Data Management for BI: Fueling the Analytical Engine*; January 2011

Author: Nathaniel Rowe, Research Analyst, Enterprise Data Management (nathaniel.rowe@aberdeen.com)