SWIM, DON'T SINK IN THE DATA LAKE: WINNING WITH BIG DATA ANALYTICS

October 2016

→ **Michael Lock,** Vice President & Principal Analyst, Analytics & Business Intelligence

Report Highlights

The deluge of data that companies face today creates a problem for those mired in an old-school technology mindset and an opportunity for those willing to embrace the diversity and flexibility of modern solutions. Companies that stay competitive in this new landscape have built a business intelligence (BI) environment that is flexible, scalable, and efficient in the collection of data and creation of insight. This report explores several Best-in-Class strategies and technologies instrumental in delivering enhanced business performance.

ABERDEEN GROUP

2

Companies prioritizing BI through the efficient use of data have a world of potential game-changing insight at their fingertips, but the first hurdle they must overcome is the "silo effect".

Research in Context

Based on findings from Aberdeen's **Big Data 2016 survey**, this report examines several Best-in-Class strategies for data management. In the context of this research, Best-in-Class companies were defined by their performance against the following three metrics.

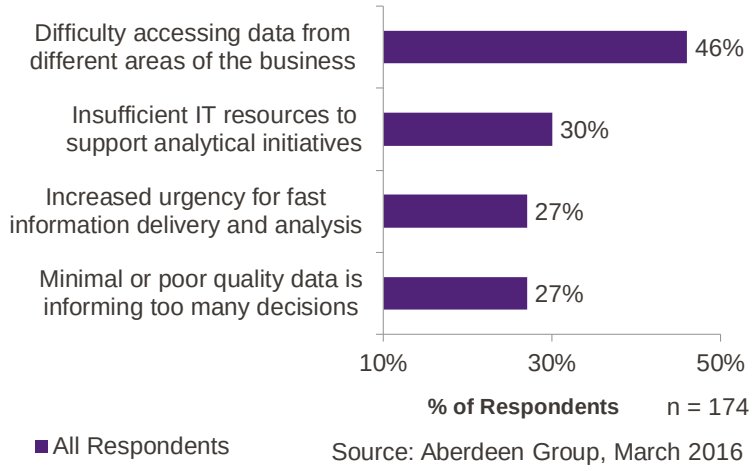
- Percentage of data records that are accurate and complete
- Percentage of respondents satisfied with data quality
- Percentage of respondents that saw reduced time spent searching for and preparing data

Drowning in Data or Surviving and Thriving?

The job of today's IT leaders has never been more challenging or ripe with opportunity. Aberdeen's research demonstrates a growing thirst for analytics and data-driven decisions among a more diverse line-of-business audience than ever before. The growing appetite for analytics bodes well for the health of these companies but adds additional strain on those charged with maintaining the data infrastructure. It's hard to ignore the fact that there is simply more data now, but the real difficulty comes in navigating the breadth of data types available. Unstructured text-based data from social channels, streaming machine data from the Internet of Things, rich media files, and geospatial information all hold tremendous potential for business insight.

Companies prioritizing BI through data efficiency have a world of potential game-changing insight at their fingertips, but the first hurdle many must overcome is the "silo effect". According to the research, accessing data from these disparate silos is the top data-related challenge that companies face today (Figure 1).

Figure I: Data Disparity Plagues Today's Organizations



Source: Aberdeen Group, March 2016

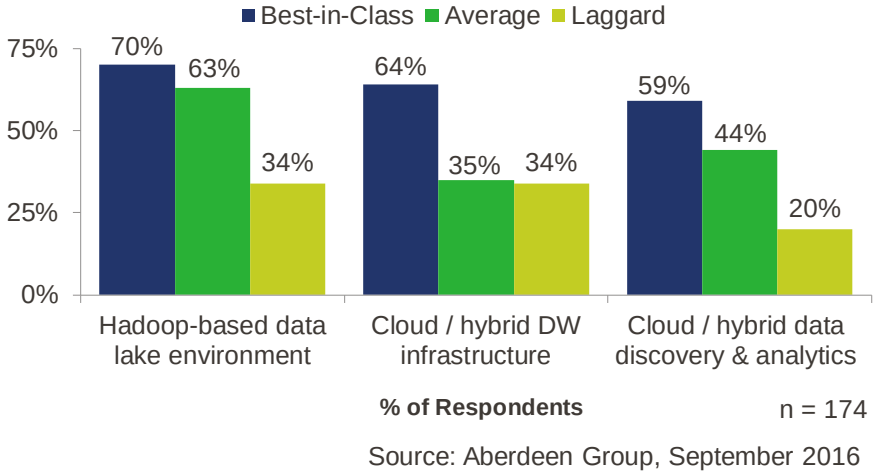
3

The other key challenges are all interrelated to a certain extent. There are more business users demanding data to support their most critical decisions, but all too often that data is incomplete, corrupted, or otherwise dirty. Confounding all these issues is a fairly noticeable and widespread lack of technical resources. In most cases companies are dealing with both a lack of manpower and breadth of expertise required manage these environments.

The Cream of the BI Crop: Top Companies Are Flexible and Agile

As companies struggle to manage this tangled web of complex data, some shrink and others face it head on. Top companies incorporate a combination of internal competency and effective use of technology (both legacy infrastructure and modern tools). On the infrastructure side, these companies look to build data flexibility and agility. A big part of that flexibility comes from their usage of Hadoop and Big Data technologies at a greater rate than Average and Laggard companies (Figure 2).

Figure 2: Exploiting the Power of Hadoop and the Cloud



Source: Aberdeen Group, September 2016

Hadoop is highly instrumental in a leader's ability to handle disparate data types simultaneously as the architecture allows for the comingling of data natively, structured or unstructured.

Traditional vs. Modern

Legacy IT infrastructure serving a variety of traditional business applications (ERP, CRM, etc.) lends itself to a traditional **enterprise data warehouse (EDW)** approach. Using relational database technology, companies can manage and organize data from a variety of structured sources.

The challenge that so many organizations face today, is that their data comes in all different shapes and sizes and doesn't mesh well with traditional RDBMS technology. Moreover, the massive and rapid growth of this environment makes scaling the infrastructure cost prohibitive.

These challenges are ultimately what drive many companies toward a **Data Lake.** Supported by a Hadoop-based technology environment, a data lake offers a repository for disparate data types comingled together in their native formats, and including all relevant data attributes. The environment also allows for distributed computing, fault tolerance, and fast scalability on commodity hardware to help stem the cost of managing data growth.

4

➔ Related Research, "The Horsepower of Hadoop: Fast and Flexible Insight with Results"

➔ Related Research "The Hybrid Data Warehouse: Fluid, Flexible, and Formidable"

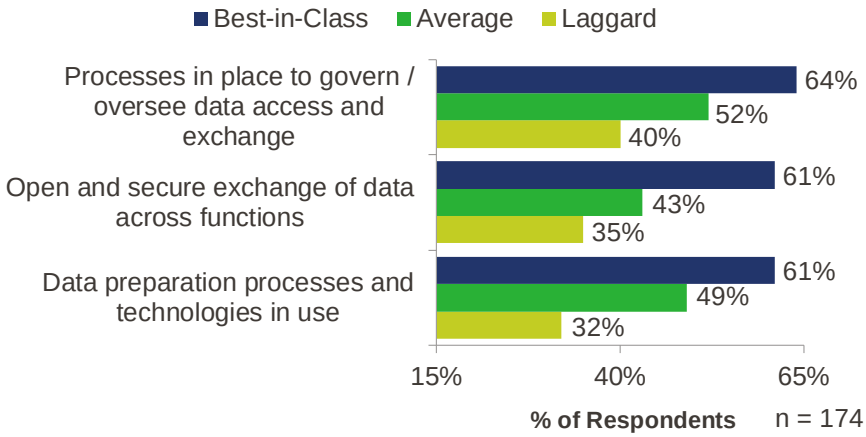
These companies are also able to maintain legacy on-premise infrastructure and applications while taking advantage of newer cloud-based technologies. These companies use a hybrid architecture in both their data warehouse environment, as well as with the analytical technologies they use. At the end of the day, top companies recognize that effective analytics is the tip of the spear when generating insight and the activities they undertake in the data layer are designed to produce that elevated level of analytical activity on the front end.

Paving the Way to Performance

The business efficiencies and financial performance that come from effective analytics are enticing, but don't arise in a vacuum. Aside from their maturity in the technology infrastructure, Best-in-Class companies are also more likely to implement critical internal processes and procedures to help ensure the cleanliness, security, and usability of their data. The stark reality of today's business and IT landscape – data breaches, phishing attacks, and good old fashioned corrupted data – necessitates strong oversight and governance. Outside of the security software and protocols in place, Best-in-Class companies employ strong data governance policies to help control access to data (Figure 3).

5

Figure 3: Twice as Likely to Share Data Cross-Functionally



Source: Aberdeen Group, September 2016

With the explosion in IoT data, unstructured social media data, and a variety of other sources, companies need a common metadata layer to centralize their understanding of what data they have and how it can be exploited.

Some might consider such policies to be red tape and the enemy of efficiency. On the contrary though, companies use governance and oversight as the foundation of quality and reliability in the data environment that actually accelerates the decision process with less wasted time searching for correct information or filling in missing fields. Once these policies are in place, top companies are more easily able to exchange information across business functions and improve accessibility to different data sources for key decision makers.

The Need for Common Metadata

More to the point of efficiency, prior Aberdeen research also demonstrates that Best-in-Class companies exploit capabilities for metadata management more frequently than their laggard counterparts. Metadata management essentially refers to the framework of information that describes the structure, location, and types of data within the organization; data vocabulary so to speak. Once again, the challenges of big data are just as much, if not more, about complexity and variety as they are about volume. With the explosion in IoT data, unstructured social media data, streaming, and a variety of other sources,

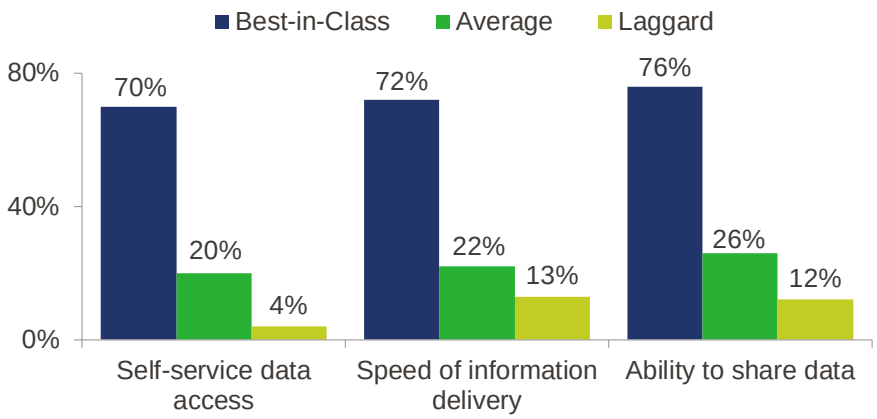
6

This self-service data environment helps support faster conversion of raw data, in a multitude of formats, into usable and consumable information, helping to quicken the pace of insight in these organizations.

companies need a way to centralize their understanding, mapping and sharing of what data they have and how it can be used. A common metadata layer should help streamline and accelerate the process of data integration and ingestion. Companies are rewarded with faster, more consistent delivery of information to key stakeholders across the business.

Ultimately, this fluidity and efficiency in managing big data pays off with results. If you think about the progression of data at a more granular level, it makes sense. Top companies have greater flexibility and scalability in their technology infrastructure which allows them to utilize more data, from more sources, and make it more accessible to a greater portion of the workforce. Security and governance policies help improve the quality and usability of data as do the data preparation technologies and activities in use. This creates an environment where more employees in more functional areas and more job roles have access to the data they need to inform their most critical decisions, ultimately contributing to a higher degree of user satisfaction (Figure 4).

Figure 4: Self-Service Data Access Absolutely Critical



% of Respondents "Satisfied" or "Very Satisfied" with above n = 174

Source: Aberdeen Group, September 2016

More Self-Service, Less IT Burden

Easily accessible and cleaner data supports a higher degree of self-service data access and less reliance on IT for every data-related need. This self-service data environment helps support faster conversion of raw data, in a multitude of formats, into usable and consumable information, helping to quicken the pace of insight in these organizations. Best-in-Class companies are more likely to get the information they need, when they need it, and before their effective decision window closes and the opportunity at hand is missed. Additionally, top companies are more satisfied with their ability to share information across functional areas. Rarely does a critical business decision depend on the insight and perspective of just one person in the organization, and Best-in-Class companies exploit efficiency in the data environment to help foster better collaboration.

Today's more successful companies are learning that the best approach is leveraging innovative technology to exploit the data as it resides in the infrastructure, essentially in-Hadoop analytics.

The Stakes Are Raised: Big Data, Bigger Performance

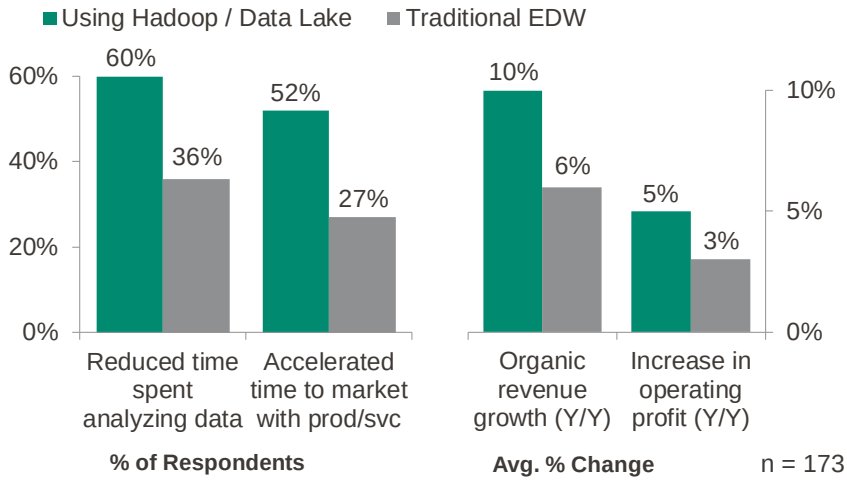
As discussed, leading companies are more likely to use innovative technologies like Hadoop and other Big Data solutions. Focusing on companies utilizing a data lake built on Hadoop, some interesting findings emerge. For starters, these companies exhibit greater agility in converting data to insight. They spend less time searching for, integrating, and prepping data, and make decisions faster. Part of this efficiency is born out of their flexibility in the decision process. Some companies initially treat Hadoop as they would a standalone data warehouse, moving data in and pulling it out as needed. However, today's more successful companies are learning that the best approach is leveraging innovative technology to exploit the data as it resides in the infrastructure, essentially in-Hadoop analytics. This approach effectively reduces failure points (and time) in the analytical process, by leaving the data where it lands, and querying against it directly in the data lake, ultimately

improving decision efficiency and boosting performance. Research shows (Figure 5) that companies using this approach were able to leverage their data agility to accelerate time-to-market.

➔ Related Research, "Running Lean Analytics with a Cloud or Hybrid Approach"

➔ Related Research "Nimble IT and the Data Layer: The Lynchpin of Analytical Success"

Figure 5: Data Agility and Business Execution



Source: Aberdeen Group, May 2016

Time-to-market however isn't the only important performance differentiator for Hadoop and other big data users. For these types of technologies and solutions, proving a tangible ROI can be difficult. However, the typical flow of value in the analytical process is certainly applicable here as well. Companies leveraging Hadoop directly for analytics and BI deliver more access to more data and make it consumable and relevant in a shorter time frame. This efficiency leads to a broader array of data-driven decisions, allowing these companies to identify and act on more opportunities for business growth and efficiency, producing substantial improvements in some critical financial metrics like revenue growth and profitability.

9

Conclusion and Recommendations

In some circles these days, the term "big data" can cause a rolling of the eyes due to overuse of the term, but not because the concepts and associated activities aren't still relevant. On the contrary, an approach to surviving and thriving in today's environment of rapidly growing and increasingly complex data has become table stakes. The research continually demonstrates a strong correlation between the ability to gain agility and efficiency in the data environment and driving enhanced business results. This is not an accident or a coincidence.

Data is the lifeblood of the modern enterprise and those who shift their worldview on big data from obstacle to opportunity will be rewarded with cleaner and more consumable data, more expedient decisions, and superior business performance. Companies considering, planning, or implementing a formal big data strategy should bear in mind the following recommendations:

- **Make self-service analytics a priority.** Linking BI with big data starts with the users. Companies need to think about ways to reduce reliance on IT in the process of accessing data in order to produce real business value. The most successful data-driven companies today are managing Big Data as the foundation of their analytical activity, as opposed to storing it and sitting on it out of regulatory requirement or corporate mandate. Moreover, any organization that deals with data as a significant aspect of their core operations (a characteristic of just about every modern enterprise) is also likely managing a heightened appetite for new data from more applications and sources. Marketers are looking to incorporate unstructured data from social channels into their customer analyses. Operations managers need better

Having to rely too heavily on IT to gather this wide variety of data puts organizations at a disadvantage. Companies need to break down unnecessary barriers to information exchange and empower their users with self-service data access.

access to streaming information from sensors and devices. Having to rely too heavily on IT to gather this wide variety of data puts organizations at a disadvantage. Companies need to break down unnecessary barriers to information exchange and empower their users with self-service data access.

➔ **Don't be stuck in the past, strategize your technology future**. Some newer, fast growing companies have been able to catch the wave of big data technologies like Hadoop and Spark just as they have become more mature and widely adopted. Many others however struggle to manage an older architecture built on legacy hardware and software systems. In relatively few cases is a complete "rip and replace" methodology appropriate. By and large most organizations should take a complementary approach and look to extract value out of prior investments while simultaneously taking advantage of newer Big Data and Hadoop-based solutions. One of the most important aspects of data flexibility is the ability to comingle and analyze traditional data with newer types of information. The tools and services are there to help companies navigate a hybrid environment that utilizes traditional on-premise data warehouse systems with newer and more flexible data architectures such as Hadoop.

➔ **Take action now, don't risk being left behind.** On one hand, efforts and investments to improve and enhance the data environment need to be well-planned and executed judiciously. On the other hand, it can be easy to fall into the trap of analysis paralysis when considering options. Taking the plunge into a more formal big data strategy is certainly not risk free, but the risk of inaction is

arguably greater. The way that we understand our businesses – improve internal operations, enhance customer experiences, identify and act on growth opportunities – rests on our ability to transform raw data into consumable insight. In a highly competitive business environment, the opportunity cost of failing to exploit this growing wealth of information is substantial, to say nothing of the actual cost and inherent security risk borne out of a poorly managed data environment. Companies should take action now to evaluate their data and analytics infrastructure and take appropriate steps to make necessary improvements and enhancements.

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

Related Research

The Horsepower of Hadoop: Fast and Flexible Insight with Results; May 2016

The Hybrid Data Warehouse: Fluid, Flexible, and Formidable; May 2016

Running Lean Analytics with a Cloud or Hybrid Approach; April 2016

Nimble IT and the Data Layer: The Lynchpin of Analytical Success; June 2015

Author: Michael Lock, Vice President & Principal Analyst, Analytics & Business Intelligence (michael.lock@aberdeen.com)

About Aberdeen Group

Since 1988, Aberdeen Group has published research that helps businesses worldwide improve their performance. Our analysts derive fact-based, vendor-agnostic insights from a proprietary analytical framework, which identifies Best-in-Class organizations from primary research conducted with industry practitioners. The resulting research content is used by hundreds of thousands of business professionals to drive smarter decision-making and improve business strategy. Aberdeen Group is headquartered in Boston, MA.