Big Data for Small Budgets

Massive volumes of business data collected from multiple sources can provide a tremendous opportunity to understand business trends and customer behavior. Though there are such clear benefits, the common assumption is that the tools and expertise needed to take full advantage of this Big Data phenomenon are restricted to the few fortunate organizations with enormous datacenters and extensive resources. However, the "Big" aspect of Big Data is often in the eye of the beholder; for smaller companies, terabytes or even gigabytes of data can pose the same problems and opportunities that petabytes do for global enterprises. Aberdeen’s Big Data research (August 2012) shows that the same tools and techniques used by these large enterprises are being adopted by smaller companies to provide fast, scalable data analytics. This report will highlight four key technologies being used to drive business value from complex data sets, and an illustrative array of affordable solutions designed for startups, SMBs, marketing departments, business units, and even cash-strapped enterprise-level IT departments.

Business Context: The Value of Big Data

The value that a Big Data initiative can bring to an organization usually falls into two major categories. First, many companies are struggling with the rapid growth of their data, which Aberdeen’s research has shown to be increasing at 40% year-over-year. This growth puts incredible strain on existing IT infrastructures and budgets, and the resulting pain points are among the most common drivers for starting a Big Data initiative (Figure 1).

Figure 1: Business Drivers for Big Data Initiatives

Source: Aberdeen Group, January 2012
IT departments struggling to collect different types of data from multiple sources and still deliver it quickly to business users are often faced with very expensive options for improving performance. These can include powerful databases, analytic appliances, advanced servers, additional data storage, network bandwidth, and all the related costs of power, space, maintenance, and headcount. Of course, a fully implemented, flexible, scalable Big Data project can alleviate many, if not all, of the above pressures, but the associated costs can seem prohibitive for smaller organizations. Fifty-one percent (51%) of small companies listed the expense of software and services as the top inhibitor preventing them from investing in improving their data environment.

The second part of the Big Data value proposition is the new business opportunities that it can provide. By successfully analyzing information on customer behavior, social media trends, product development, sales and marketing effectiveness, or any of a thousand other data types, new insights can be uncovered to help steer the business. Organizations that implemented a Big Data initiative reported a 12% year-over-year improvement in their operating profit, and grew their total customer base by over 14% in the last fiscal year. These performance improvements are more than 26% better than companies without these initiatives.

Given the economic difficulties that persist in much of the world today, most organizations are looking at ways to both address their IT struggles and capture new data-driven business opportunities without breaking the bank. There are four technological areas critical to Big Data success, each with affordable options available:

- **Data integration**
- **In-memory computing**
- **Unstructured data management**
- **Data visualization**

### Integrating Disparate Data Sources

One of the most common business pressures driving Big Data investment is having too many data silos (Figure 1 above). In many companies, important business data is spread out over too many locations, from databases to file stores to collaborative web portals to multiple versions of enterprise applications like ERP or CRM systems. Big Data companies had on average 20 unique internal data sources that stored data necessary for operations or analysis.

However, Big Data environments don’t end at the edge of a corporate firewall. In the modern global, interconnected business environment, being able to quickly and efficiently send and receive data from business partners is vital. On average, Big Data organizations conducted data exchanges with over 12 different partners, suppliers, distributors, data aggregators, or OEM resellers. Furthermore, these companies added valuable data from nine

---

Definitions: Maturity Class

Organizations were classified according to their performance in the following metrics:

**Leaders** in Big Data reported

- √ **83%** success rate for delivering information quickly enough to meet user demand
- √ **5 hours** per week per knowledge worker spent searching for information
- √ **35%** of all active data immediately accessible for queries / analytics

**Followers** in Big Data reported

- √ **61%** success rate for delivering information quickly enough to meet user demand
- √ **10 hours** per week per knowledge worker spent searching for information
- √ **13%** of all active data immediately accessible for queries / analytics
other external sources. Figure 2 shows the most popular sources were internet generated, such as social media, web traffic, and customer sentiment. However, these sources can also include economic information, stock prices, customer purchase behavior, census data, postal codes, or even weather data.

**Figure 2: Key Data Sources Used in Big Data Initiatives**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>All Respondents (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transactional / Structured</td>
<td>93%</td>
</tr>
<tr>
<td>Social Media / Customer Sentiment</td>
<td>86%</td>
</tr>
<tr>
<td>Clickstream</td>
<td>83%</td>
</tr>
<tr>
<td>Internal Unstructured</td>
<td>82%</td>
</tr>
<tr>
<td>Human Resources</td>
<td>77%</td>
</tr>
<tr>
<td>Location / Geo-spatial</td>
<td>67%</td>
</tr>
<tr>
<td>Digital Media / Rich Media</td>
<td>61%</td>
</tr>
<tr>
<td>M2M / Sensor</td>
<td>60%</td>
</tr>
</tbody>
</table>

Source: Aberdeen Group, January 2012

Given the complexity inherent in accessing data from so many sources, an important piece of a Big Data foundation is the ability to easily move data from one source to another. Three quarters (74%) of Leaders in Big Data had **ETL tools** (extract, transform, load) to move their data, and they were 1.6-times more likely than Followers to be able to integrate data in real time.

**Processing Big Data at High Speed with In-memory**

The most common business pressure, as shown in Figure 1, is an inability to deliver information as quickly as business users need it. Efficiently delivering data and analysis on business events as they occur, and making informed decisions on this information, are marks of an agile, data-driven organization. The demand for data is growing, and business users are asking for it faster than ever. Forty-seven percent (47%) of Big Data organizations need insight within an hour of a business event occurring, and more than a third (35%) need it in near real time.

One of the most effective technologies for processing data at high speed is in-memory computing. These solutions load the target data directly into the random access memory (RAM) of a server or desktop, very close to the processor itself. This eliminates the need to connect to a storage array or disk, locate the desired information, and convey it over a network to the server doing the processing. Without these potential bottlenecks, the full

---

**Fast Facts: Demand for Data**

Business users are asking for faster and faster access to business data and insight.

- **40% of Large Companies** needed actionable intelligence within **one hour** of a business event.
- **28% of Large Companies** needed actionable intelligence at **near real-time** speeds.

Small organizations, however, had almost exactly the same demand for data:

- **40% of Small Companies** needed actionable intelligence within **one hour** of a business event.
- **29% of Small Companies** needed actionable intelligence at **near real-time** speeds.

“Address the quality and consistency of your data at the source or as close to the source as you can. Pay attention to consistent business processes and to naming standards with properly identified global meaning across the enterprise.”

~ Manager

$1B - $2.5 Billion North American Chemicals Company
power of the processors can be directly used to access and manipulate the desired information.

Organizations that have adopted this technology experience substantial improvements in their data operations (Table 1). These companies tend to have larger data volumes to start with, but they are able to analyze a much larger percentage of their data, and do so at speeds over two orders of magnitude faster than their competitors.

### Table 1: More Data, More Speed, More Efficiency

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Use in-memory computing (n = 33)</th>
<th>Don’t use (n = 163)</th>
<th>In-memory Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median amount of active business data</td>
<td>38 terabytes</td>
<td>18 terabytes</td>
<td><strong>2.1 times more data</strong></td>
</tr>
<tr>
<td>Median amount of data analyzed</td>
<td>14 terabytes (37% of all data)</td>
<td>4 terabytes (22% of all data)</td>
<td><strong>3.5 times more data</strong></td>
</tr>
<tr>
<td>Average response time for data analysis or query</td>
<td>42 seconds</td>
<td>75 minutes</td>
<td><strong>107 times faster</strong></td>
</tr>
</tbody>
</table>

Source: Aberdeen Group, January 2012

There are several in-memory solutions aimed almost exclusively at large global enterprises. However, the basic concept behind the technology can translate from advanced next-generation servers to machines as simple as a standard notebook. Given the steadily decreasing cost of RAM, most commodity servers have dozens or hundreds of gigabytes of memory available, opening the door for low-cost solutions.

### Handling the Unstructured Data Problem

One of the major strengths of Big Data initiatives is the ability to collect, manage, and analyze not just structured data from relational databases, but unstructured or semi-structured data from documents, emails, social media feeds, images, video, and rich media. A surprising amount of business data resides in these unstructured formats; Aberdeen’s February 2012 research on *Handling Paper in a Digital Age: The Impact of Document Management* put the overall amount at 50% of all data.

There are a number of different database management systems being designed to leverage unstructured information. On a broad level, most are referred to as NoSQL databases, commonly interpreted as "not only structured query language.” This departure from traditional relational database management systems (RDBMS) gives more flexibility to the data formats being stored, managed, and accessed, as well as serving as an important repository of data to feed into analytic platforms.
One of the most exciting developments in the world of unstructured data analysis is Apache Hadoop, an open-source file storage framework that couples the flexibility of managing unstructured data with high-powered processing capabilities. Based on Google's MapReduce programming model for breaking large, complex problems into small bite-sized chunks, Hadoop can use the processing power of clusters of ordinary servers to tackle tasks far larger than any single machine could accomplish. Given the fact that the code is open source, and can run on commodity hardware, it offers a powerful, cost-effective Big Data option. For more information on this technology, see Aberdeen's report, *The Little Elephant in the Big Data World: Hadoop 1.0 Goes Live* (March 2012).

**Figure 3: Awareness for Hadoop**

![Figure 3: Awareness for Hadoop](image)

Hadoop is still a fairly new technology, and has flown under the radar of many organizations. Over two-thirds (69%) of business users either haven't heard of it or don't understand what it does. Those in the IT community have a more developed understanding, but still close to half (45%) are oblivious to this opportunity. However, for those organizations that have heard the message, this technology has become one of the hottest areas for investment. Aberdeen's research shows that companies are reporting the adoption rate of Hadoop and similar MapReduce technology will more than quadruple in the next 24 months.

**Making Sense of Big Data with Visualization**

The final step in any analytics project is to take the processed information and make it easily digestible and accessible for its intended audience. This involves the visual rendering of business insight, often in the form of reports, dashboards, or charts. Aberdeen's July 2012 research into *Agile Business Intelligence* reported on the continuing trend of moving away from static reports managed by IT and towards more self-service, interactive solutions put in the hands of the business user.

This trend is even more pronounced in Big Data environments, for the simple fact that the scale and complexity of the data infrastructure puts a
premium on IT resources, and makes it prohibitively inefficient for them to continue to manage BI reporting. A further benefit of investing in solutions that are easy to use for the line of business is that employees with the most expertise in business operations are the ones exploring the data, increasing the likelihood of discovering important nuggets of insight.

**Figure 4: Understanding and Visualizing Data**

As Figure 4 shows, the Leaders in Big Data implementation are far more likely to have implemented both traditional and interactive methods for accessing and digesting business information. With interactive reports and dashboards, not only is the data summarized and presented through charts and graphs, but these visualization elements can be dynamically updated as the user changes parameters or drills down into subsets of the information to find underlying patterns or relationships.

However, not all data analysis can be accomplished through easy interactive interfaces. Advanced data techniques such as predictive analytics, or advanced technologies like Hadoop, usually require a particular skill set. Many organizations realize the necessity of employing people who are solely dedicated to supporting, exploring, and mining the mountain of business data being collected. These data scientist or business analyst roles can sift through the dross and share valuable information with others in the organization to use or act upon. Almost two thirds (63%) of Leaders employ someone in that dedicated role, compared to only two out of five (42%) of the Followers. Even small companies are targeting these highly skilled positions, with 66% reporting either having filled or trying to fill this type of job role.

**Summary and Key Takeaways**

Big Data initiatives can help organizations alleviate some of their IT pain points, and open up new doors for business insight. For organizations or

"Big data is a huge opportunity to gather insight into what’s going on in the enterprise’s ecosystem, especially customers. We need the technology to analyze and get the results we are asking for."

~ IT Manager,
Large EMEA
Software Company
departments looking to take advantage of these opportunities without a lot of resources, Aberdeen suggests evaluating four key technologies that have affordable options:

- **Data integration tools** allow organizations to quickly and easily move data from one source to another. Big Data companies draw from over 41 unique data sources, and much of the value of Big Data comes from new and interesting mashups. So it is no wonder that Leaders outstrip Followers in adopting this technology.

- **In-memory computing** provides companies with a method for processing data at high speed by moving the information directly into the RAM of a machine. Companies with this technology get data results over 100-times faster than companies without.

- **Unstructured data management tools** are ways organizations can handle the information that doesn’t fit neatly into a traditional relational database. There is a lot of potential business value hidden in these data types, which is why technologies such as Hadoop are projected to grow more than four times in the next two years.

- **Data visualization tools** and self-service BI solutions allow for business users and decision makers to easily explore and digest business information. This final layer of an analytic stack builds on top of the powerful hardware discussed above, and makes insight available to everyone in the organization. Again, Leaders in Big Data are up to 1.7-times more likely than Followers to have implemented these tools.
For more information on this or other research topics, please visit www.aberdeen.com.

<table>
<thead>
<tr>
<th>Related Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Go Big or Go Home? Maximizing the Value of Analytics and Big Data;</strong> September 2012</td>
</tr>
<tr>
<td><strong>The State of Big Data: Video Benchmark;</strong> July 2012</td>
</tr>
<tr>
<td><strong>Agile or Fragile? Your Analytics, Your Choice;</strong> July 2012</td>
</tr>
<tr>
<td><strong>Beyond Agile Analytics: Is Agile Data Integration Next;</strong> June 2012</td>
</tr>
<tr>
<td><strong>Managing the TCO of BI: The Path to ROI is Paved with Adoption;</strong> May 2012</td>
</tr>
<tr>
<td><strong>Enabling Access to Big Data with Data Integration;</strong> April 2012</td>
</tr>
<tr>
<td><strong>High Performance Organizations Empower Employees with Real-Time Mobile Analytics;</strong> April 2012</td>
</tr>
<tr>
<td><strong>Mobile BI 2012: Accelerating Business on the Move;</strong> March 2012</td>
</tr>
<tr>
<td><strong>The Little Elephant in the Big Data World; Hadoop 1.0 Goes Live;</strong> March 2012</td>
</tr>
<tr>
<td><strong>Divide and Conquer: Using Predictive Analytics to Segment, Target and Optimize Marketing;</strong> February 2012</td>
</tr>
<tr>
<td><strong>Operational Intelligence - Part 1: Driving Performance with Tactical Visibility;</strong> February 2012</td>
</tr>
<tr>
<td><strong>In-memory Computing: Lifting the Burden of Big Data;</strong> January 2012</td>
</tr>
<tr>
<td><strong>The Role of Big Data Analytics in HR: Speed, Satisfaction and Scale;</strong> January 2012</td>
</tr>
<tr>
<td><strong>Agile BI: Complementing Traditional BI to Address the Shrinking Decision-Window;</strong> November 2011</td>
</tr>
<tr>
<td><strong>Agile BI: Three Steps to Analytic Heaven;</strong> March 2011</td>
</tr>
</tbody>
</table>

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