In Big Data We Trust?

Businesses Are Hungry For Big Data Insights, But Trust Issues Erode Confidence

April 2013
Table Of Contents

Executive Summary .......................................................................................................................... 2
Tired Old Platforms Cannot Keep Up .......................................................................................... 2
A Few Succeed While Others Struggle ......................................................................................... 5
Technology Matters — Start With A Platform You Can Trust .................................................. 7
Key Recommendations: ................................................................................................................ 7
Appendix A: Methodology ............................................................................................................ 8
Appendix B: Supplemental Material ............................................................................................ 8
Appendix D: Endnotes .................................................................................................................. 8

© 2013, Forrester Research, Inc. All rights reserved. Unauthorized reproduction is strictly prohibited. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change. Forrester®, Technographics®, Forrester Wave, RoleView, TechRadar, and Total Economic Impact are trademarks of Forrester Research, Inc. All other trademarks are the property of their respective companies. For additional information, go to www.forrester.com. [1-KY5R3J]
Executive Summary

Fueled by the buzz and a few public success stories, many firms are counting on big data to deliver tomorrow’s critical business insights. While the hype is deafening, we think reality offers both promise and obstacles, including a perception that emerging big data technology is not capable of supporting mission-critical, real-time analytic applications — the exact thing firms need for success. In August 2012, Intel commissioned Forrester Consulting to evaluate this hypothesis and provide recommendations to IT professionals on their options.

We analyzed data from several Forrsights surveys and conducted in-depth interviews with six senior IT executives with knowledge of their firm’s analytic requirements and plans. We spoke both with firms having success with big data and those that were still looking for it; we found that firms need to overcome business and IT misalignment issues and build solutions on trustworthy platforms in order to be ready for secure, real-time, production-grade solutions.

Key Findings

Forrester’s study yielded three key findings:

- **Firms need more real-time insight from more data.** Firms are investing in business intelligence and decision support solutions to get insight from an increasing amount of digital data, but they realize that yesterday’s technologies do not meet tomorrow’s need for real-time, predictive analysis at extremes of scale.

- **Firms are dissatisfied with their current technology capabilities.** Our study found firms struggling to implement affordable solutions with their current technology stack. A few successful firms were turning to big data technologies such as Hadoop, but they took different approaches to ensuring that the platform was capable of meeting their needs.

- **Firms must create a big data platform capable of supporting secure, reliable production-grade solutions.** In order to succeed beyond proofs of concept, pilots, and science experiments, firms deciding to pursue big data with Hadoop need a platform capable of enterprise-class production operations. The selected Hadoop distribution must support both high-latency batch and robust, iterative, real-time analysis on the same cluster in a reliable fashion with a comprehensive security approach that includes fine-grain access control and encryption.

Tired Old Platforms Cannot Keep Up

The explosion of data cannot positively impact your business unless you have the ability to use insights gleaned in time to make a difference. Big data technology such as Hadoop has received a lot of attention lately for its potential to deliver deep insights from data at extreme scale, but are firms ready to deploy it for enterprise use? Forrester interviewed six business technology executives with knowledge of their firm’s analytic and data investment plans and authority over execution. In addition, we examined several Forrsights business technology survey data products looking for insights.

We found that expectations for big data are high, but reality has created some steep hurdles to success. Specifically firms are:
Turning to data with big expectations for value. According to our Forrsights Hardware Survey, 48% of companies maintain between 100 TB and 9 PB of data in total. The top drivers for data growth are business expansion, more data from each business activity, and increased compliance requirements (see Figure 1.) One executive we spoke with was piloting a big data project, and he expected 300% growth in the year after the project went live. Another related the explosion of human genomic data being made available through research and government mandate.

We found it telling that despite coming out of a recession and IT budget’s being tight, firms keep expanding the data they collect and plan to use. Clearly, data is where the money is in the emerging information economy.

Expecting shiny, new BI tools to yield real-time insight. Our Forrsights Software Survey indicates that firms are planning to increase use of business intelligence and decision support technology (see Figure 2). Our interviews, however, reveal a deeper need — firms must be able to make future predictions on real-time, raw data that is creating a deluge of information that their current technology cannot handle. For example, a director of IT of a large shipping company told us:

“We have an ongoing problem. We aren’t getting timely information from our terminals, and this is costing us money. We are starting to instrument containers with GPS, but this creates a deluge of data that we then need the systems to analyze.”

Firms that expect new BI technology to deliver what they need are destined for disappointment; batch loaded data in rigid data warehouses can’t deliver the predictive, real-time insight required.

Seeing the potential in integrating non-traditional data. Most executives interviewed wanted to peer into the combination of transactional, structured and non-traditional unstructured data to pick needles out of the haystack. For example, an IT manager for a large hardware and software solutions manufacturer told us:

“In order to react to trends, we must munge through petabytes of data, and that is an issue . . . we need a way to park TBs of unformatted, unstructured data and do analytics across the structured data and unstructured.”

Technology for extreme-scale batch data analysis is maturing; however, it is often not suitable for enterprise-grade applications. For example, many Hadoop vendors are making big strides in high availability architectures; however, the Apache specification for this is very young and immature. Furthermore, we find that query performance across petabytes of unstructured data prevents near-real-time iterative analysis that businesses need.

Setting a blinding pace. During our interviews, the pressure executives feel was palpable — use of more data, more rapidly to make better decisions was a business imperative. A VP of global threat intelligence at a large information security firm told us:

“Our desire is to increase the pace at which we can see the connections between data of different structures and formats that is growing in different silos. We need much quicker analytics in order to see what is happening closer to real time and make decisions.”

Firms acquiring new BI packages and implementing big data solutions such as Hadoop have a tall order to fill — meet the need for real-time, enterprise-grade analytic solutions despite the risk associated with immature technology.
Forrester Consulting
In Big Data We Trust?

**Figure 1**
Business Growth Is The Top Reason Data Is Growing

"Why is the amount of data stored by your firm increasing?" (Please select the top three reasons)

- Business is growing; increased data generated by existing systems: 46%
- We are capturing more data per business activity than before: 32%
- Regulatory compliance and auditing requirements: 29%
- Replication and disaster recovery: 28%
- Data warehousing, reports, and analytics: 26%
- Increased audio and video data: 22%
- We simply save everything: 21%
- We just don't want to throw anything away: 21%
- Longer retention period for key business data/backups: 16%
- Increased PC backups: 12%

Base: 1,281 IT decision-makers
Source: Forrsights Hardware Survey, Q3 2012, Forrester Research, Inc.

**Figure 2**
Business Intelligence And Analytics Are The No. 2 Priority

"Which of the following initiatives are likely to be your IT organization’s top software priorities over the next 12 months?"

- Update/modernize key legacy applications: Critical priority 21%, High priority 41%
- Increase our use of BI and analytics: Critical priority 17%, High priority 44%
- Upgrade packaged applications to a newer release: Critical priority 15%, High priority 46%
- Consolidate or rationalize enterprise applications: Critical priority 14%, High priority 41%
- Invest in mobile apps on smartphone or tablets: Critical priority 13%, High priority 35%
- Increase deployment and use of collaboration technologies: Critical priority 11%, High priority 43%
- Increase our use of SaaS: Critical priority 10%, High priority 30%
- Increase our use of IaaS: Critical priority 9%, High priority 27%
- Expand use of Agile software development and processes: Critical priority 8%, High priority 23%
- Outsource application support and maintenance: Critical priority 5%, High priority 15%
- Reduce the number of (major) software vendors: Critical priority 8%, High priority 19%
- Increase our use of open source: Critical priority 3%, High priority 10%

Base: 1,204 IT decision-makers
Source: Forrsights Software Survey, Q4 2011, Forrester Research, Inc.
A Few Succeed While Others Struggle

We found that some firms successfully use more data for deeper insight, while others struggle. For example, we found a global shipping company using real-time data to predict profitability before it agrees to carry cargo, and a genome analytics company handling trillions of rows of data and billions of correlations. The leaders we talked to at these firms had confidence in their abilities and were willing to invest in advanced analytic solutions such as complex event processing and Hadoop. But leaders at other firms did not reflect such confidence. We found:

- **Across the board agreement that old school techniques are not cutting it.** A common theme in our interviews was the inability of current IT systems to deal with the scale of data required. In one case, a director of IT admitted to attempting to implement a large-scale data integration and analytics solution on a relational database technology that failed because the indexes were not adequate and could not be rebuilt in a cost-effective manner. Now they are pursuing Hadoop to solve the problem, but that brings its own concerns about suitability for enterprise applications.

- **A general distrust of data caused by years of underperformance and high costs.** According to our Forrsights Software Survey, improving business decisions is the top expectation of business intelligence, analytics, and decision support tools but surprisingly, number two is improvement in data quality and consistency (see Figure 3). This reflects an underlying issue of trust in the data and the systems that handle it. Our interviews underscored our assessment — attempting to deal with the types, volume, and velocity of data needed today with yesterday’s technology creates a further erosion of trust in the data. Will it be good enough? Can I rely on it? Will systems that I invest in be stable and give our business what it needs? These were all common concerns.

**A hesitancy to dive into “big data.”** While IT leaders at firms are prioritizing investments in BI and analytics, they are unsure about “big data” technologies (see Figure 4). This is not surprising considering the hype. On the other hand, we found one firm that made a big commitment to skilling up for Hadoop because the IT department felt that their core business value proposition depended on it. Another was moving to a supported Hadoop distribution because the IT leaders there felt that they needed better reliability and runtime support. The execs we spoke with at a third firm were concerned about their internal ability to operate Hadoop and were turning to a partner for operational support in the near term. Each of these firms implemented a different strategy that allowed it to face big data technology uncertainty with confidence. Simply buying new BI tools like Hadoop and capturing more data won’t do it. Confidence is the biggest differentiator between firms that succeed in getting more insight from more data and those that do not. Confidence comes from trust.
**Figure 3**
Improve Data Quality Is The No. 2 Reason Firms Invest In Business Intelligence

“What changes, if any, has your firm experienced or is your firm expecting over the next 12 months from an increased use of BI, analytics, and decision-support tools and services?"  

<table>
<thead>
<tr>
<th>Change Expectation</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make better informed business decisions</td>
<td>72%</td>
</tr>
<tr>
<td>Improve data quality and consistency</td>
<td>69%</td>
</tr>
<tr>
<td>Improve business planning and execution</td>
<td>64%</td>
</tr>
<tr>
<td>Monitor and improve business process performance</td>
<td>63%</td>
</tr>
<tr>
<td>Improve customer satisfaction</td>
<td>56%</td>
</tr>
<tr>
<td>Invest in dashboard capabilities</td>
<td>53%</td>
</tr>
<tr>
<td>Provide overall competitive advantage</td>
<td>51%</td>
</tr>
<tr>
<td>Build internal skills for required BI technologies</td>
<td>43%</td>
</tr>
<tr>
<td>Improve compliance and risk management</td>
<td>42%</td>
</tr>
<tr>
<td>Increase budget to implement BI and decision-support solutions</td>
<td>26%</td>
</tr>
<tr>
<td>Shift budget from more traditional back-office application to BI solutions</td>
<td>18%</td>
</tr>
<tr>
<td>Hire third-party services to implement BI solutions</td>
<td>15%</td>
</tr>
<tr>
<td>Increase the number of IT users</td>
<td>14%</td>
</tr>
<tr>
<td>Improve collaboration with business partners using the same SaaS apps</td>
<td>13%</td>
</tr>
<tr>
<td>Don’t know</td>
<td>1%</td>
</tr>
<tr>
<td>None of the above</td>
<td>1%</td>
</tr>
<tr>
<td>Other</td>
<td>0%</td>
</tr>
</tbody>
</table>

Base: 1,204 IT decision-makers  
Source: Forrsights Hardware Survey, Q3 2012, Forrester Research, Inc.

**Figure 4**
Firms Are Uncertain About Big Data

“What are your firm’s plans to adopt the following business intelligence and analytics technologies?"  
(Respondents who answered “Don’t know"

<table>
<thead>
<tr>
<th>Technology</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytics based on in-memory technology</td>
<td>13%</td>
</tr>
<tr>
<td>In-process analytics (e.g., Hadoop, BigTable)</td>
<td>13%</td>
</tr>
<tr>
<td>Complex event processing</td>
<td>11%</td>
</tr>
<tr>
<td>Text analytics</td>
<td>10%</td>
</tr>
<tr>
<td>Decision-support solutions</td>
<td>7%</td>
</tr>
<tr>
<td>Business performance management solutions</td>
<td>7%</td>
</tr>
<tr>
<td>Data quality management</td>
<td>7%</td>
</tr>
<tr>
<td>Advanced analytics (e.g., predictive analytics)</td>
<td>7%</td>
</tr>
<tr>
<td>Social media analytics</td>
<td>7%</td>
</tr>
<tr>
<td>Specialized database engines</td>
<td>6%</td>
</tr>
<tr>
<td>Web analytics</td>
<td>5%</td>
</tr>
<tr>
<td>Business intelligence solutions for mobile apps</td>
<td>5%</td>
</tr>
<tr>
<td>Data visualization, dashboards</td>
<td>4%</td>
</tr>
<tr>
<td>Reporting tools</td>
<td>3%</td>
</tr>
</tbody>
</table>

Base: 1,332 IT decision-makers  
Source: Forrsights Software Survey, Q4 2011, Forrester Research, Inc.
Technology Matters — Start With A Platform You Can Trust

Successful firms approach big data as a business/technology partnership in which the business provides the analytics skills and IT provides the platform. We agree with this approach and advise IT managers to consider the technology platform’s trustworthiness in light of current and future performance needs. The best big data platform:

- **Does not deplete the resource bank.** Big data technology can be tricky, and your team must learn new tricks. For example, having separate storage, server, and network infrastructure teams makes managing a cluster of networked commodity computers challenging. Further, there are almost 200 command lines required to tune Hadoop File Store (HDFS). Firms need a solution that can be reliably maintained with the resources they have, not one that requires a different organization and a completely new set of skills.

- **Stands up to real real-time workloads.** When we asked one executive what kept him up at night about his big data investment, he was concerned that the solution would be so successful that his business would come to depend on it, but if it broke he would not be able to fix it. Big data platforms must evolve beyond unreliable and arcane batch analytics into real-time uses that are easy to operate and maintain. Firms must think through future analytics requirements, especially those that require robust, dependable responsiveness. Cheap or free now rarely meets tomorrow’s needs — managers must plan for real-time, extreme-scale analytics today.

- **Engenders confidence that data is secure.** Security of unstructured data is a perennial top concern of executives thinking about investing in big data analytics, and our interviews supported this trend. Firms are concerned that big data systems collect huge volumes of data that has not been classified, analyzed, or protected by the same sophisticated encryption and field-level access controls available in traditional database technology. In order for a big data platform to be trusted for production solutions, it must go beyond simple file security by providing data profiling, record-level access control, and logging capabilities.

**KEY RECOMMENDATIONS:**

- **Decide where “big data” fits into your firm’s core value proposition.** Clearly some companies make their money on exploiting extreme-scale data, such as the genomics firm we interviewed. Others such as the shipping company had many real-time big data needs. Others had just a few. Figure out where you are on this continuum before deciding how much risk to take and how much support you need when implementing a big data platform.

- **Improve the quality of analytics by iterating on raw, big data in real time.** Traditional predictive analytics models rely on small, cleansed data sets that can be significantly different from actual data collected in operations. Big data offers the techniques and technologies to develop predictive models that use larger, fresher data. Since the data can be raw and dirty, iterative solutions yield the best results. For example, a firm may decide to run several petabyte-scale data transformations using the Hadoop Pig language before using Mahout to iterate through machine learning models at extreme scale.

- **Look for enterprise-ready Hadoop platforms.** For those deciding to embrace Hadoop for big data, select a distribution that is capable of handling real-time, iterative workloads in a production environment. Unfortunately, many of the Hadoop vendors have a long way to go here. However, we expect that to change in the next 18 months as vendors are rushing to shore up their Hadoop offerings to run in a true enterprise production environment.
Appendix A: Methodology

In this study, Forrester conducted qualitative surveys of 10 IT decision-makers in the United States to evaluate their data management practices. Questions provided to the participants asked about current and future data management practices and perceptions. Respondents were offered a small cash incentive as a thank you for time spent on the survey. The study began in August 2012 and ended in October 2012.

Interviewed:

1. VP of global threat analytics at a large information security software vendor.
2. IT manager for a large hardware and software solutions manufacturer.
3. CEO of genome analytics service provider.
4. Director of IT for state government agency.
5. Chief technology officer for a large manufacturing company
6. Director of IT for a global shipping company

Appendix B: Supplemental Material

Related Forrester Research


Appendix D: Endnotes

1 Each calendar year, Forrester’s Forrsights for Business Technology fields business-to-business technology studies in more than 17 countries spanning North America, Latin America, Europe, and developed and emerging Asia. For quality control, we carefully screen respondents according to job title and function. Forrester’s Forrsights for Business Technology ensures that the final survey population contains only those with significant involvement in the planning, funding, and purchasing of IT products and services. Additionally, we set quotas for company size (number of employees) and industry as a means of controlling the data distribution and establishing alignment with IT spend calculated by Forrester analysts. Forrsights uses only superior data sources and advanced data-cleaning techniques to ensure the highest data quality. We have illustrated only a portion of survey results in this document. For access to the
Forrester Consulting

In Big Data We Trust?

full data results, please contact forrsights@forrester.com. Forrsights surveys cited: Forrsights Hardware Survey, Q3 2012; Forrsights Software Survey, Q4 2011; Forrsights Budgets And Priorities Tracker Survey, Q2 2012.

2 According to our June, 2011 Global Online Big Data Survey of 60 IT professionals, 70% said that big data projects were business/IT collaborations.