

TURNING BIG DATA INTO BIG DOLLARS

With Cloud Computing And Open Source Software

Written By Anand Bhadouria
Sr. IT Strategist, Enterprise Cloud Solutions

Table of Contents

1. Executive Summary	2
2. What is Big Data, anyway?	3
3. Big Data is big—and getting bigger	3
4. Big Data promises many benefits	4
5. But using Big Data is challenging	5
6. The solution: Move Big Data to the cloud...	7
7. ... and use open source software	8
8. Finding a partner to help	9
9. Notes	10

1. Executive Summary

The future of your enterprise could be at stake if you don't start turning Big Data into big dollars.

Big Data is the flood of digital information streaming in through everything from social media to embedded sensors. It's too big to handle with traditional tools, and is growing at the astounding rate of 40% to 60% a year.

Any enterprise can benefit from Big Data. In fact, research shows that a typical Fortune 1000 company that uses data only 10% more effectively can generate \$2 billion in added revenue. In the consumer space, retailers can increase margins by 60%. The U.S. healthcare system can save a shocking \$200 billion a year. With existing competitors and new startups jostling for position, every enterprise must pay attention to Big Data.

But there are big challenges including incompatibilities with existing infrastructure, high capital expenses and the threat of vendor lock-in. Further compounding the challenge, the strategic skills needed to use Big Data are in short supply.

To generate cost-effective results from Big Data, enterprises need to look beyond on-premises IT to cloud computing. The cloud is perfectly suited to Big Data with massive storage, elastic expandability, and pay-as-you-go billing. Instead of maintaining your own costly infrastructure, with cloud computing you pay just for what you use when you need it.

By working with a proven technology partner, most enterprises can save time and trouble as they look to turn Big Data into big dollars. Rackspace is an ideal technology partner, with a broad portfolio of offerings, open source software for building public and private clouds, and a stellar reputation for professional services.

"The future of your enterprise could be at stake if you don't pay attention to Big Data"

2. What is Big Data, anyway?

Big Data refers to the continuous stream of digital information that arrives through many channels, including:

- e-mail
- social media
- user-generated content like photos and videos
- customer and B2B transactions
- GPS-equipped cell phones
- mobile devices
- system logs of all kinds
- RFID tag sensors embedded in everything from airport runways to casino chips

This data is 80 to 90 percent unstructured, meaning it doesn't have any predictable format like a pro-forma report. And there's too much of it for an enterprise to handle with the same tools and expertise used for routine business.

What's more, Big Data is a moving and rapidly growing target, not pegged to any fixed number of gigabytes or terabytes. As more data accumulates in a given sector, tools must scale up to handle it, so the goal posts continuously move further out.

3. Big Data is big—and getting bigger

Information today flows relentlessly in all directions, and accumulates faster than ever before. Massive volumes of data are now common place in most organizations, in every sector, in every country. In fact, *The Economist* calls data "the new raw material of business: an economic input almost on a par with capital and labor."¹

Big Data is growing at a ferocious rate of 40 to 60% a year.

Almost any organization, in any sector, can benefit from Big Data.

Most people are familiar with megabytes and gigabytes, but not the larger terms used to discuss Big Data. Datasets are now being measured in terms of terabytes, petabytes, and exabytes — exotic terms that aren't even

in most spell checkers. See Table 1 for definitions and examples to help visualize them.

These terms are now essential because the world's stock of information is growing at a ferocious rate of 40% to 60% a year, depending on whose estimates you prefer.²

The bottom line? Big Data is with us today, and will keep growing for the foreseeable future.

Table 1: Terms Used for Big Data Storage Capacity		
Term	Size	Example of Capacity
Gigabyte (GB)	1,000,000,000 (one billion) bytes	1 GB = 2 hours of CD-quality audio or 7 minutes of HDTV
Terabyte (TB)	1,000,000,000,000 (one trillion) bytes	1 TB = 2,000 hours of CD-quality audio, or almost 5 days of HDTV
Petabyte (PB)	1,000,000,000,000,000 (one quadrillion) bytes	1 PB = 7 weeks of HDTV, or 1.5 million 64 GB iPods full
Exabyte (EB)	1,000,000,000,000,000,000 (one quintillion) bytes	1 EB = 16 months of HDTV, or 15 million 64 GB iPods full

4. Big Data promises many benefits

Almost any organization in any sector can benefit from Big Data. Major retailers like Walmart and eBay use Big Data collected from transactions to refine their online search engines and encourage customers to buy more.^{3,4}

Food and beverage suppliers like Pepsi's Latin American division merged customer, logistics, and manufacturing data to significantly improve its plant operations.⁵

Even non-profits like Children's Hospital Los Angeles use leading-edge computer science to search pediatric datasets and find patterns that can help doctors save infant's lives.⁶

These examples touch on the most promising benefits of Big Data:

- Better productivity
- Higher revenues
- Better decision-making
- Innovative new products, services, and business models

"Big Data was once thought of as a tool for competitive advantage, but in today's competitive landscape, it is essential for survival," adds Ryan Hawk, Rackspace VP of Information Management.

Some attempts have been made to quantify these benefits. A study at the University of Texas found that if a median Fortune 1000 business increased the usability of its data by only 10%, it would earn an extra \$55,900 per employee — or \$2+ billion a year in added revenue.⁷

If the typical Fortune 1000 business uses data 10% more effectively, it can boost revenues by \$2+ billion a year

McKinsey Global Institute says retailers could increase their margins 60% by squeezing more value from their Big Data. And the U.S. healthcare system could save a shocking \$200 billion a year.⁸

On the flip side, companies that ignore Big Data are vulnerable to competitors. If you miss this wave, you risk being left behind by nimbler competitors. Or as Forrester puts it in a recent report, “Your firm’s future depends on effectively using more data.”

Every organization in the world needs to take Big Data seriously.

Some key sectors positioned to benefit strongly from Big Data include computers and electronics, entertainment and media, finance, healthcare, information services, insurance, retail, and all levels of government. All these sectors generate a tremendous amount of data that can be mined for insights or repackaged with new business models.

Beyond these, enterprises in every sector can benefit in marketing, product development, and customer service. With so much to gain, every organization in the world needs to take Big Data seriously.

5. But using Big Data is challenging

Management teams are often expected to “do more with less.” But in this case, they’re being asked to “do more with more”...more Data that is. To generate useful insights, Big Data must be acquired, stored, and analyzed in thoughtful and cost-effective ways.

But there can be many challenges to using Big Data, such as:

- Existing infrastructure can’t handle Big Data
- Higher capital expenses (CapEx)
- Higher costs for non-strategic resources
- Vendor lock-in that limits choices

EXISTING INFRASTRUCTURE CAN'T HANDLE BIG DATA

Because Big Data is less structured than information filed away in any CRM or ERP system, traditional databases and analytics platforms cannot handle it. In fact, Big Data is unexplored terrain that will stretch and strain any company's IT infrastructure.

"Capturing, filtering, storing and analyzing Big Data flows can swamp traditional networks, storage arrays and relational database platforms," notes a recent article from MIT Sloan. "Attempts to replicate and scale the existing technology will not keep up with Big Data demands."¹⁰

HIGHER CAPITAL EXPENSES

If an IT team tries to handle Big Data without rethinking an existing on-premise infrastructure, this will usually mean buying rack after rack of drives to accommodate growing data, and adding scores of servers for occasional peaks of processing demand. All this can lead to sky-rocketing and unpredictable capital expenses, or CapEx, which most CFOs like to avoid, especially for rapidly growing data.

HIGHER COSTS FOR NON-STRATEGIC RESOURCES

The more servers and storage drives brought in-house, the more IT personnel needed to tend them. These added resources fill non-strategic roles doing little besides routine maintenance, or simply "keeping the lights on." They will not contribute the kind of strategic insights that Big Data analysts could.

VENDOR LOCK-IN LIMITS CHOICES

The issue of vendor lock-in can result from buying proprietary hardware or software that doesn't work with other products. That limits any future options for updating your infrastructure or switching suppliers.

When you're locked into one vendor, you have to keep buying from them, even if you never see an acceptable ROI from their wares. You're caught in a vicious circle.

"Legacy systems and incompatible standards and formats too often prevent the integration of data and the more sophisticated analytics that create value from Big Data," notes McKinsey.¹¹

In the worst-case scenario, an enterprise can be locked into a costly and inflexible infrastructure that can't even support the analytics tools needed to delve into the company's Big Data.

When you're locked into a vendor with a closed business model, you're stuck buying from them, even if you never see an acceptable ROI.

LOOMING SHORTAGE OF SPECIALIZED SKILLS

Another huge challenge is the looming shortage of people who can help enterprises take advantage of Big Data. This new trend increases the demand for skills such as data acquisition, analysis, and visualization. Anyone with these skills can now write their own ticket in one of hottest new roles in IT: a Data Scientist.

It's going to get harder for businesses to find and retain experienced Data Scientists to meet the demand for deep, Big Data analysis.

These lateral thinkers are responsible for sifting out answers to mission-critical questions like, "Who are our customers?" and "Why do they buy from us?" and "How can we get them to spend more with us, more often?"

Rather like Mr. Spock in the original Star Trek, this role is a "hybrid of data hacker, analyst, communicator, and trusted advisor," says the Harvard Business Journal, calling it "the sexiest job of the 21st century." But as the key players in Big Data, the shortage of data scientists is becoming "a serious constraint" and "competition for top talent will remain fierce."¹²

Cloud computing is well-suited to Big Data, and can provide better services plus a stronger bottom line

By 2015, Big Data demand will reach 4.4 million jobs globally, but only one-third of those jobs will be filled, says Gartner.¹³ The reason? Colleges have fallen behind, and just can't equip graduates with the right skills fast enough. Enterprises will have to pay dearly to hire from the small pool of qualified data scientists, plus offer incentives like signing bonuses and stock options.

For all these reasons, your existing IT team and infrastructure may not be well-equipped to handle any Big Data projects. Your management team should look beyond your current situation for a more effective solution.

6. The solution: Move Big Data to the cloud...

To squeeze the most value from Big Data at the least cost, and move beyond the limitations of in-house IT, many enterprises are turning to the cloud.

"Cloud computing is well-suited to Big Data," notes MIT Sloan, since the cloud matches the need for cost-effective mass storage, elastic expandability, and pay-as-you-go pricing.¹⁴

For example, a company might need more computing resources to handle a burst of Big Data analytics on an unpredictable schedule. The cloud provides a simple and affordable way to spin up as many extra servers as needed, release them as soon as the job is complete, and pay only for the time they were actually used.

Most CFOs consider the utility billing for cloud computing, paid out of operating budgets or OpEx, to be better for the corporate balance sheet. This means the cloud can provide the best of both worlds: better services plus a stronger bottom line.

Forrester researcher Holger Kisker recently named three key reasons why using Big Data in the cloud makes sense:

1. Big Data requires a spectrum of advanced technologies, skills, and investments. Do you really want this all in-house?
2. Big Data includes huge amounts of external data. Does it make sense to move and manage all this data behind your firewall?
3. Big Data needs a lot of data services. Why not focus on the value of your analysis, instead of simply managing your data?¹⁵

For all these reasons, moving your Big Data projects to the cloud makes a lot of sense.

7. ... and use open source software

Here's the key decision to make when moving to the cloud: Should your enterprise use proprietary commercial software or open source?

Much of the web and the cloud already run smoothly on open source software like the LAMP stack (an acronym for Linux, Apache, MySQL, and Perl). Probably the most popular tool for handling Big Data, Hadoop, is award-winning open source software.

For maximum flexibility and cost-effectiveness in the cloud, hundreds of enterprises are now looking at the open source software called OpenStack.[®] This is a set of flexible, cost-effective software tools that make it faster and easier to build public or private clouds. Since it's open source — available under the Apache 2.0 license — there's no purchase fee and absolutely no threat of vendor lock-in.

WHAT IS OPENSTACK?

OpenStack is an open and scalable operating system for building public and private clouds. It provides both large and small organizations an alternative to closed cloud environments, reducing the risks of lock-in associated with proprietary platforms. OpenStack offers flexibility and choice through a highly engaged community of over 6,000 individuals and over 190 companies.

Rackspace offers a broad portfolio of products and services for OpenStack cloud platforms.

- Powerful and scalable public cloud platform
- Private cloud, anywhere— in your data center or ours
- Hands-on training so you can handle your own OpenStack clouds

8. Finding a partner to help

Enterprises need a technology partner who can help them turn Big Data into big opportunity. With so much at stake — even company survival — what should an enterprise look for in this partner?

It makes sense to seek out a partner who can offer deep expertise in cloud computing, a portfolio of open source software, a proven commitment to support, and seasoned professional services. It is imperative that a partner not only be able to design and run an optimized architectural environment, but also have access to the expertise needed for Big Data analytics.

IT budgets and internal technologies will not scale to handle the amount of infrastructure and human resources necessary to handle current and future big data needs. That is why the only logical path to conquering big data is leveraging the benefits of cloud computing and finding the right software and expertise to sift insight from petabytes of data and operationalize those insights into action.

Rackspace is well suited to partner with enterprises to ensure they have a means to mine the wealth of knowledge in their big data while avoiding strain on already pinched budgets and expertise.

Seek out a technology partner that offers deep expertise in the cloud, a portfolio of open source software, a proven commitment to support, and seasoned professional services

Rackspace's Enterprise Cloud Solutions organization with its team of Sr. IT Strategists and Cloud Solutions Architects are acknowledged leaders in all of these areas and Rackspace's strategic alignment with Hortonworks, a leader in Apache Hadoop development, implementation, support, operations and training, provides customers with an enterprise-ready Hadoop platform that is easy to use in the Cloud. Together, Rackspace and Hortonworks focus on eliminating the complexities and time-consuming, manual processes that are required for implementing Big Data solutions. The joint effort pursues an OpenStack-based Hadoop solution for public and private cloud, which can easily be deployed in minutes. Private cloud solutions are available now with a public cloud solution scheduled for release soon.

Let Rackspace's team of Enterprise Cloud Solution experts teach you how technology can help you to crunch, sort and use Big Data so that it is valuable to your business.

Schedule your complimentary Advisory Services: IT Evolution Workshop today. Together, we will define how a Big Data solution can strengthen your business, construct your optimized environment and create an actionable roadmap to get you there.

Contact us at 1-800-440-1249 or send us an email at advisory_services@rackspace.com

9. Notes

1. ["Data, data everywhere," The Economist, 25 February 2010](#)
2. Different sources estimate Big Data growth differently. The Economist and several analysts says digital information increases 10x every five years, equal to 60% annual growth. IBM says data is growing at 50% a year. IDC predicts the digital universe will grow at more than 45% between 2010 and 2015, while sales of related technology will grow by 40% a year.
3. ["Report: Wal-Mart's Big Data Moves Will Boost Rackspace," Data Center Knowledge, 22 October 2012](#)
4. ["How Ebay uses big data to make you buy more," ZDNet, 20 October 2012.](#)
5. ["How Big Data Came to PepsiCo," Forbes, 12 December 2012.](#)
6. ["Using big data to save lives," Phys.org – Computer Sciences, 22 October 2012.](#)
7. ["Measuring the Business Impacts of Effective Data, chapter 1," University of Texas, September 2010, p3](#)
8. [James Manyika et al, "Big Data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, May 2011, p64, p39](#)
9. [Brian Hopkins and Boris Evelson, "Expand Your Digital Horizon With Big Data", Forrester, 30 September 2011, p2](#)
10. [Thomas Davenport, Paul Barth and Randy Bean, "How 'Big Data' is Different," MIT Sloan Management Review, 30 July 2012](#)
11. [James Manyika et al, "Big Data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, May 2011, p12](#)
12. [Thomas H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, October 2012](#)
13. ["Gartner Reveals Top Predictions for IT Organizations and Users for 2013 and Beyond," press release, Gartner, 24 October 2012](#)
14. [Thomas Davenport, Paul Barth and Randy Bean, "How 'Big Data' is Different," MIT Sloan Management Review, 30 July 2012](#)
15. [Holger Kisker, "Big Data Meets Cloud," For Chief Information Officers Forrester blog, 15 August 15 2012](#)

ABOUT THE AUTHOR

Anand Bhadouria is a Sr. IT Strategist, with Rackspace Enterprise Cloud Solutions, who helps enterprise customers build public cloud, private cloud and hybrid infrastructures that deliver optimal benefits from Big Data. He has more than 20 years experience working with Fortune 500 enterprises including banks, insurance firms, manufacturers, pharmaceutical companies and retailers. Anand received his Master of Science degree in Management and Engineering from the MIT Sloan School of Management and MIT College of Engineering. Anand can be reached at anand.bhadouria@rackspace.com.

ABOUT RACKSPACE

Rackspace® Hosting (NYSE: RAX) is the open cloud company, delivering open technologies and powering more than 190,000 customers worldwide. Rackspace provides its renowned **Fanatical Support**® across a broad portfolio of IT products, including public cloud, private cloud, hybrid hosting and dedicated hosting. The company offers choice, flexibility and freedom from vendor lock-in. Rackspace has been recognized by Bloomberg BusinessWeek as a Top 100 Performing Technology Company and is featured on Fortune's list of 100 Best Companies to Work For. Rackspace was positioned in the Leaders Quadrant by Gartner Inc. in the "2011 Magic Quadrant for Managed Hosting." Rackspace is headquartered in San Antonio, TX with offices and data centers around the world. For more information, visit www.rackspace.com.

DISCLAIMER

© 2013 Rackspace US, Inc. All rights reserved.

This White Paper is for informational purposes only. The information contained in this White Paper is selected from public sources which we believe is reasonable in relation to subject matter and are believed to be accurate. RACKSPACE MAKES NO WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, AS TO THE INFORMATION OR SPECIFIC PRODUCTS. Users must take full responsibility for application of any product mentioned herein.

Rackspace and Fanatical Support are service marks of Rackspace US, Inc. registered in the United States and other countries. OpenStack is a trademark or registered trademark of OpenStack, LLC in the United States and/or other countries. All other product names and trademarks used in this document are for identification purposes only to refer to either the entities claiming the marks and names or their products, and are property of their respective owners. We do not intend our use or display of other companies' tradenames, trademarks, or service marks to imply a relationship with, or endorsement or sponsorship of us by, these other companies.