4sythcomm.com

White Paper

For Big Data Analytics There's No Such Thing as Too Big

The Compelling Economics and Technology of Big Data Computing

March 2012

By: 4syth.com
Emerging big data thought leaders

Table of Contents

Executive Summary

Facebook, Twitter, Google, and Yahoo, are companies that have always dealt with big data. Now, as ever-more data deluges organizations globally everyone else has to know how to handle it. That's because it's not going away since a lot of machine-generated data – from sensors, weblogs, imagery and data streaming from devices – is growing along with Moore's Law. The more devices generating data out there, the more data piles up in the data center. While this may cause headaches initially because it's more complicated, it brings huge new benefits. Among these, according to one important study, is that companies taking advantage of the superabundance of data through "data-directed decision-making" enjoy an up to 6 percent productivity boost. In another insight, this from the IBM 2010 Global CFO Study, over the next three years, organizations that leverage big data will financially outperform their peers by 20 percent or more. Erik Brynjolfsson, professor of management science at Massachusetts Institute of Technology's Sloan School of Management, says, "it's not just big data in the sense that we have lots of data. You can also think of it as 'nano' data, in the sense that we have very, very fine-grained data – an ability to measure things much more precisely than in the past. You can learn about the preferences of an individual customer and personalize your offerings for that particular customer." The only way to keep up with the escalating data is to expand beyond traditional RDBMS tools in a way that most enterprises have never done. The tools to do that are only coming into business today. Hadoop, an open source product, is a favorite among developers who are working with big data. Since it is open source, the cost is minimal, not including hardware and support. Cisco and Oracle have delivered the first-ever enterprise class NoSQL solution for harnessing vast volumes of real-time unstructured and semi structured big data. It comes with software from Oracle in the form of Oracle NoSQL software and innovative hardware from Cisco Unified Computing System and enterprise class support from both titans in the technology industry.

The State of the Industry

Here's the problem: recall when a terabyte of data was hard to get. No longer. Today, we're deluged with data – it's expanding beyond terabytes into petabytes, and even exabytes (1 million terabytes). So much so that some organizations with limited storage systems are telling analysts they're throwing terabytes away. Ask someone who knows and they'll say those enterprises are trashing metaphoric gold.

Not everyone – not even a majority – will accept this until more go through the data rethink tank, meantime there's the competitive advantage early adopters are having now.

That's because the data in their terabytes and petabytes is providing answers to questions like: "What are the social networks saying about our reputation?" or "How is the blizzard impacting traffic now?" or "What are influential bloggers saying about what colors and fabrics are growing in popularity, and who might be able to supply us those materials now?"

The ability to get specific answers to such questions – and many more – comes with the growth of what's called big data. In the last decade, organizations have gone from just accumulating standard data to nonstandard forms; from static to dynamic in a paradigm shift that's altered thinking about enterprise data assets. Innovative software has taken command to capture all this and provide new insights in organizations that are accessing big data, leading to faster, better decisions and quicker responses.

Data is entering systems at a rate that follows Moore's Law, doubling every two years. In two days we see as much data as we did from the dawn of civilization through 2003, according to academic studies. That's every two days. Twitter alone is closing on almost 100 million tweets every 24 hours. This data flooding from customer loyalty programs, remote sensor technology, call centers and social media combined with decades of collecting data in data warehouses can be enormously valuable.

"If I'm right about data being generated not by people but by machines talking to machines, Moore's Law is creating new data at a furious pace," says Michael Olson, founder and CEO of Cloudera, which markets a new open-source data platform called Hadoop to handle big data.

Yet studies show that a majority of organizations aren't sure how to maximize the demonstrable value. Some business leaders see only the mounting costs of storing so much data, not the value in it. The big data imperative is not to discard it but to use it to begin an analytical journey leading to better insight all round. "Businesses that are getting all of these status updates (on Facebook) and user-generated messages today need to understand all this and how to digest it," says Olson.

This paper aims to explain what's changing today as we advance information by including everything in real or near real-time, and to describe what's behind the emerging next generation of technology that's augmenting the familiar traditional relational database management systems (RDBMS) when they can't handle big data sets in a cost-effective and timely manner.

Research company IDC believes that between now and 2020, the amount of information in the digital universe will grow by 35 trillion gigabytes (1 gigabyte equivalent to 40 (four-drawer) file cabinets of text, or two music CDs). In 2011 alone, the amount of digital information created and replicated surpassed 1.8 trillion gigabytes – growing by a factor of nine in just five years, according to IDC. That's on par with the number of stars in the physical universe.

To put another perspective on this, in 2015 global mobile data traffic flow over the Internet will reach 6.30 exabytes a month, according to a Cisco estimate. One exabyte is the equivalent of 10 billion copies of a regular printed weekly news magazine.

Enterprises Need Robust Big Data Solutions

As the name implies, big data connotes something that is enterprise level. It follows that enterprises need a robust, commercially supported hardware-software solution to deal with it. Today's big data innovations came from Web 2.0 companies producing a growing collection of open source technologies that changed the culture of collaborative software development, and the scale and economics of hardware infrastructure. These technologies enable data storage management and analysis in ways that were not possible before with more costly traditional technologies, such as traditional RDBMS.

NoSQL is one such technology that has emerged as an increasingly important part of Big Data trends and the broader data store landscape. NoSQL often is characterized by what it is not, and industry definitions vary. It can be not only an SQL-based or simply not an SQL-based relational database management system. NoSQL databases form a broad class of non-relational database management systems that are evolving rapidly, and several solutions are emerging with highly variable feature sets and few standards.

As interesting as these technologies are for all the innovations they bring, not all solutions meet enterprise application requirements. Many organizations require dependable, supported, and tested enterprise-class solutions for rapid deployment of mission-critical applications.

To address these needs, Cisco and Oracle are the first vendors to collaborate to deliver integrated, enterprise-class NoSQL solutions. Enterprises cannot afford the cost of developing and maintaining a big data solution that relies on an "army of developers" and unreliable hardware. A complete hardware-software integrated solution for addressing big data needs is necessary for them. Cisco and Oracle provide that solution. Exceptional performance, scalability, availability, and manageability are made possible by the Cisco Unified Computing System (UCS) and Oracle NoSQL Database. Together, this solution provides a platform for the quick deployment along with predictable throughout and latency for most demanding applications.

Best of all, this solution blends with existing Cisco UCS infrastructure for Oracle Database installations, and offers enterprise-class service and support from both companies, dramatically reducing the risk and time associated with deploying NoSQL solutions.

A recent Oracle-Cisco joint big data white paper is available at www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns944/le_34301_wp.PDF.

Big Data: Are You Ready For It?

Where's all this growth coming from? In two ways: "structured" (standard), meaning organized by codes, classifications, or common practice in a standard form that computers can read. Examples are real estate documentation that identifies the house owner, lists the address where mortgages can be foreclosed; debts, taxes and rates collected; deliveries made; and where utilities can control services and collect their bills, and the NAIC code (North American Industry Classification System standard for classifying businesses). In other words, structured data is identifiable in a fixed field within a record or file. The rest is "unstructured" (nonstandard) or semi-structured data. This include blogs on social networking sites, geolocation devices, bar codes, vehicle telematics, stock market trades, x-rays, phone conversations, videos, text messages, contracts, images, ads, spreadsheets, delivery directions, Word docs, audit trails, and e-mails. In other words, unstructured data has no identifiable structure and is, therefore, not so easily usable. Even this is changing as Web content is increasingly "tagged", and facial and voice recognition software can identify people and words in digital files.

What is Big Data?

Since the work of software scientists is to join people and all things together they call it big data. Big data is the result of a convergence of trends. Technology to collect, manage and store data today is way cheaper (1/6th the cost of six years ago, according to IDC), and dramatically quicker than ever; data is viewed as really valuable – especially understanding customer behavior, or improving customer satisfaction, or increasing traffic flow – and a third trend is understanding that gut feel alone in today's complicated world is simply not the way to run organizations, big or small.

"Traditionally, the term 'big data' has been used to describe the massive volumes of data analyzed by huge organizations like Google or research science projects at NASA," says Merv Adrian, research vice president at Gartner.

"But for most businesses, it's a relative term – it begs the question, 'what's big'? 'Big' depends on an organization's size," adds Adrian. "What's big is a constantly moving number and always assessed by comparison to the vast amounts some companies work with. But big data as a concept in IT parlance today, Adrian says, tends to mean something fairly specific, not just about size but also about composition and the nature of the processing." The point is more about finding new value within and outside conventional data sources. Two-thirds of the firms Tech Target surveyed are keeping more than a year's worth of data online – 43 percent have more than three years' worth.

"Pushing the boundaries uncovers new value and opportunities, and 'big' depends on where you start."

Lev Manovich, professor at the Visual Arts Department, University of California-San Diego, author of the 2008 book, *Software Takes Command*, defines big data as "data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set."

"Big data is all about being able to see data through more and finer grain context and understand your customers and your operating environment in that context," says John Akred, senior manager, Technology R&D, Accenture.

"It is daunting for many organizations to grasp the amount of big data being produced from sensors and devices, let alone getting business value from it. Yet with a platform approach to data, organizations can monitor, detect and predict changes and impact to the business by having a better handle on such data coming from connected devices," adds Akred.

Acquiring Big Data

While Cloudera's Olson is as impressed as anyone with big data, he doesn't think that big data by itself is the most important part of the coming paradigm shift. "What is really interesting is the variety of data we are getting now and the tools we want to use to analyze it," he says.

Variety is one of the four main components – and there are others – of big data, known by some as the "four Vs":

- **Volume** – enterprises are swimming in data, banks, for example collect data in terabytes (the U.S. Library of Congress' total book stack measures 15TB – and petabytes (Google processes more than 1PB every hour).

  Managing huge volumes of data is a major challenge for financial services firms, for example. Data sharing across Wall Street enterprises is still a big issue, as each business unit frequently prefers calculating from its own set of data. With these companies collecting even more unstructured data, advanced enterprises have developed tools that can analyze news, via video, audio and Twitter, for example, in real-time to help make trading decisions.

  New regulations focused on transparency and risk management to be put in place in 2012 are driving greater urgency among capital markets firms to manage big data.

  "Big data in financial services in 2012 – or 2015, or 2020 for that matter – is going to be an important topic," said James Austin, CEO of Vertex Analytics. "There aren't going to be any firms that want less data in the future; they are all going to want more data. And regulations are going to play a large part in big data," he told Wall Street & Technology magazine in November, 2011.

  Bank of America Merrill Lynch, for example, is using Hadoop technology to manage petabytes of data for advanced analytics and new regulatory requirements.
- **Velocity** – global banks handle trillions of messages in a single day's trading, mostly processed by computers.
- **Variety** – the IT industry has dealt with big data for decades as structured data in static and disciplined databases and spreadsheets.

  What's new are tools to effectively capture, visualize and analyze unstructured data that is messy, moving, ubiquitous, streaming in text, audio, video, clicks, PDF files, email, blogs, tweets, sensors and the rest. About 80 percent of a corporation's data is unstructured including office productivity documents, e-mail, Web content, in addition to social media. Email and messaging systems create unstructured data more than anything else. While two of five respondents to a Unisphere survey say upper management is barely aware of the challenges of unstructured data, IT professionals are seriously concerned about the volumes they're getting. At least 57 percent of respondents report that unstructured data is very important, and about 18 percent consider it a core of their business.
- **Variability** – semantics, or the variability of meaning in language.

Using Big Data

Use cases are being generated in various industry segments and spreading quickly into newer areas. Here are some examples:

Amazon's retail price competition:
Amazon burst into the headlines in December, 2011 with a promotion offering iPhone and Android users $5 off for sharing in-store prices while shopping for toys, electronics, sports, music, and DVDs, with their mobile phone application. The twofold aim was to increase usage of its Amazon's Price Check App bar-code scanning application while also collecting comparative intelligence on store prices. Using the Amazon app. shoppers scan a bar code, take a picture of the item or conduct a text search to find the lowest price.

Fed BP Disaster response:
Early in 2010's Deepwater Horizon disaster, the oil rate flow was a major question. BP and independent groups offered varying estimates hindering efforts to coordinate the scale and scope of the U.S. Government's response. With close enough not good enough, the National Institute of Science and Technology (NIST), worked to make sense of the disparate estimates. NIST used the open source R statistics language to analyze and reconcile the estimates and produced actionable intelligence on which to base the ultimate response.

Analyzing social media streams:
One company measures influence across the social web by storing, processing, and analyzing real time social media data streams. The company's platform analyzes signals as they travel through the social Web and performs natural language processing, machine learning, and other analysis to measure topical and broad based influence.

Cross-selling:
One company increases an average order size by recommending complementary products based on predictive analytics for cross selling.

Email:
An insurance company implements big data analytics to retain emails and other electronic documents in anticipation of future litigation and investigations. In the case of a legal hold placed on the company's electronic information, or an internal investigation the archived data is searched and delivered within the system.
The capacity to speedily access vast quantities of previously inaccessible electronic information due to its unstructured format accelerates the legal process, aids investigative analysis, and improves the odds of mounting a successful case.

Fraud:
Governments and credit card companies are detecting fraud (including claims and tax) in online systems in real-time based on behavioral patterns.

Human behavior:
Companies are able to quantify aspects of human behavior never before accessible. Social networks, news stream, and the smart grid, are way of measuring "conversation", "interest", and "activity". What's more, with data machine-learning algorithms and big data tools to reveal insight, it used only the structure, not the content, of the data. One company has collected 3 billion tweets from 60+ million users tokenized into 16 billion+ usages of 65 million terms – more than a terabyte of data. It can identify whom to follow to understand how events and news stories resonate, and even to find dates.

Keyword campaigns:
Retailers use big data analytics to drive traffic from search engines to their Web sites. The software collects information about which keywords work best to turn Internet surfers into shoppers. They also create language models so that their sites can return product results if the shopper enters a related word. For example, if a shopper searches for “dining table” on a retail site it will return results for dining room furniture. The retailer may be able to return in future to offer new styles of dining room furniture more likely to appeal based on that particular shoppers’ tweets and Facebook updates.

Location tracking:
A company collects information for its mapping service it sells to large businesses. The company can tap into data from probes and mobile devices around the world to collect traffic data. To figure out information about a particular street, the company formerly had staff poring through hundreds of terabytes of data.

Market and customer segmentation:
An arms race is occurring in the retail sector. If retailers can understand consumer behavior and collect behavioral data to better guide product decision-making, then every cent they can eke out is increasing their margins, and allowing them to invest more.

Monitoring mobile phone usage:
Usage information such as average screen brightness, signal strength, wifi connectivity, power cycles, and more is collected daily and sent to a cluster. Data is sent when the phone is connected to a power source to avoid using battery to send data upstream. Once the company’s monitor has been running for a few weeks on a phone, it starts analyzing the usage data and makes recommendations in the form of push notifications. A notification might suggest turning on auto screen brightness, or start using wifi when the signal is low.

Policing power tools:
Using big data analytics makes for more dynamic law enforcement. In North Carolina police are using a Web based application that offers courts and law enforcement agencies access to integrated criminal data. A view of an offender is seen through a single application, allowing for positive identification through a photograph as well as an “offender watch” capability to alert law enforcement professionals of an offender’s change in status. The servers integrate data within a state’s criminal justice applications.

Managing risk:
The North Carolina Department of State Treasurer is turning to analytics software to manage the state’s pension risk to better protect the pensions of more than 850,000 residents. The risk management technology, customized for state pensions, will help offices in the department’s Investment Management Division better assess the risks associated with new and current investments in its $72.8 billion portfolio.

Risk modeling:
A large bank combined separate data warehouses from multiple departments into a single global repository for analysis. The bank used the cluster to create a new, more accurate score of the risk into its client portfolios. This enables the bank to better manage its exposure, and to offer its clients better products and advice.

Tracking ads:
A critical premise for success of online advertising networks is to successfully collect, organize, analyze and use large volumes of data to place an ad appropriate to a Web page user. Given the nature of their online orientation and dynamics, it is critical that these processes be automated to the largest extent possible. Specifically, the success of advertising technology and its impact on revenue are directly proportional to its capability to use large amounts of data in order to compute proper impression value given the unique circumstances of ad serving events

such as the characteristics of the impression, the ad, and the user as well as the content and context. As a general rule, more data results in more accurate predictions.

Customer churn:
A large telecommunications company provided analyzed call logs and complex data from multiple sources. Using predictive analytics across this data it was able to predict the likelihood of any individual customer cancelling (churn). The information also helped the company create more sensitive relationships with customers to reduce churn.

Organizing Big Data

Big data is about connecting the dots of all the content that's out there by analyzing a huge data set and returning a set of results in milliseconds.

It becomes possible to scan all Web images for specific facial characteristics, for example, to link these to different ill-defined data connected to the images, and to filter and create a subset of data for additional analysis, and to link case data to identify terrorists.

The possibilities arising from this evolving ecosystem makes it clear that big data is not like your father's business intelligence (BI) tools. It could mean, for example, designing an enterprise data warehouse (EDW) to support technology that analyzes data – hence, the term analytics – to help organizations enter the big data era successfully.

Because big data crunches data sets that are so large they cannot be speedily analyzed by traditional database software tools, analytics is emerging with innovative software products – purposely designed for large amounts of data in all forms, including text, numbers, images and voice.

Ralph Kimball, founder of the data warehouse consulting company Kimball Group, says use cases like those outlined above are motivating a search for architectural similarities and differences across all of them, and more. What they seek in this fluid environment is a system architecture that addresses big data analytics generally.

"Big data is a paradigm shift in how we think about data assets, where do we collect them, how do we analyze them, and how do we monetize the insights from the analysis," says Kimball.

Because the "old ways of organizing knowledge no longer work," writes journalist-author James Gleick, in his book *The Information: A History, a Theory, a Flood*, a new approach to conventional data sources, therefore, is necessary. The fact is that the software and hardware we've relied on are not capable of capturing, managing, or processing the new forms of data within reasonable development or processing times.

As enterprises look seriously at unstructured data for the first time, asking how on earth they analyze it, Kimball sees the EDW entering an exciting new realm with a dramatically increased set of functions.

Two Architectures Emerge

Two architectures have emerged to address big data analytics: extended RDBMS, and MapReduce/Hadoop. These architectures are being implemented as completely separate systems and in various innovative hybrid combinations involving both architectures. At the same time, Cisco and Oracle have delivered the first-ever enterprise class NoSQL solution for harnessing vast volumes of real-time unstructured and semi structured big data. It comes with software from Cloudera, the leading provider of Hadoop management tools and support. Hadoop – named for child's stuffed elephant – is considered reliable because the Hadoop Distributed File System (HDFS) essentially stores, and MapReduce essentially analyzes.

James Kobelius, a senior Forrester Research analyst, sees Hadoop as the nucleus of the next-generation EDW. Hadoop implements the core features that are at the heart of most modern EDWs: cloud-facing architectures,

massively parallel processing, in-database analytics, mixed workload management, and a hybrid storage layer. MapReduce/Hadoop is a grouping of components to handle distributed data-processing including masses of unstructured data such as Twitter tweets, email, instant messages, security and application logs. Hadoop supports the MapReduce programming model developed initially by Google for efficiently processing big data and, significantly, to reduce cost of managing that data.

These architectural options provide organizations with a choice between using their traditional date warehouses and their existing data warehouse architectures, deploying Hadoop or merging the two. This third option is seen by analysts as more likely for organizations to progress beyond basic BI reporting to data mining at depth and to predictive analytics.

Hadoop is reportedly favored by some early adopters as stand alone cloud analytics infrastructure. For others such as Philip Russom, research director for data management at The Data Warehousing Institute (TDWI) it's complementary to the EDW. Organizations can certainly try to use traditional RDBMS for big data workloads, but it's considered likely to be much more expensive to build a system with equivalent performance for a given workload that's better suited to Hadoop. This being so, RDBMS and Hadoop are considered more than likely to converge in the coming years since neither alone will sew up the big data analytics market. One of Hadoop's challenges is that it comprises a half-dozen separate software pieces that need integrating to get it to work. The problem is that this requires expertise that is in short supply.

Jeff Hammerbacher, cofounder of Cloudera and a Hadoop expert, says organizations need first to see data as a competitive advantage before building a big data function. The next step is to build out a low-cost, reliable infrastructure for data collection and storage for whichever line of business they perceive to be most critical. If organizations don't have that digital asset, then they're not even in the game. Once there, they can start layering on the complex analytics. Most companies go wrong when they start with the complex analytics.

Recent architectural developments for extended RDBMS are massively parallel processing (MPP), and columnar storage. Combined, extended RDBMS is able to scale to support analysis of exabytes of big data (1 EB is equivalent to 10 million copies of a printed weekly news magazine), and other requirements. Other developments involve technologies that can be distributed over thousands of geographically scattered processors, provide millisecond responses to tightly guarded standard SQL queries, and updating data in place at full load speeds.

Despite these advances, extended RDBMS isn't a single solution for big data analytics, according to Kimball, because it can't handle more specialized requirements such as relational semantics for many of the complex use cases big data analytics requires.

More than one-third of respondents in a 2011 survey of 611 data managers and professionals report they manage most of their company's information – including all information types, such as text, video, or audio – within core enterprise databases.

These data managers from the International Oracle Users' Group surveyed by Unisphere Research say they are struggling with rapid data growth, but few have control over the storage technologies used to manage this growth. In many cases, DBAs don't have a great awareness of accumulated or projected storage costs. As data grows, the reflex reaction by most organizations is to buy and install more disk storage.

Despite the smart approaches on the horizon, they're only prevalent among a minority of companies.

Close to one-third of companies now embrace tiered storage strategies, and only one out of five is putting information life cycle strategies into place to better and more cost effectively manage their data. A number of companies are compelled to preserve data for extended periods, for example, to meet compliance requirements.

As a result, more data is being kept online for longer periods – which increases storage costs. In fact, 12 percent of respondents say they simply now hang on to all data "forever."

"Suddenly we've got all of these status updates (on Facebook) and user generated content messages that we need to understand, and now we've got to digest that stuff," says Cloudera's Olson.

It's easy to see that as devices generate data it keeps piling up and the faster and cheaper these devices get the faster and cheaper data gets and goes on piling up.

It's not the fault of enterprises having to face the situation that's responsible for this, it's just that new data is being created at such a furious rate while they're dealing with the complexity of supporting systems they can't keep up with large scale distributed data generation unless they have large scale distributed data stores.

Enterprises seeking to transcend this situation need to look at Netflix or Amazon with service oriented architectures to manage the complexity of their application infrastructures and using the new emerging technologies to deal with the data at scale and deliver it to them.

"People still use relational databases for lots and lots of work," said Olson. "They're really good if you've got predictable queries that run over pretty structured, tabular data. That's a lot of work. All of the BI and the enterprise data warehouse technology in the market is of long-term value. That's not going away. Oracle, IBM, and others have long-term markets there and are going to keep selling to it.

"What's happening now is that data people want to work with is getting more complicated," adds Olson.

The problem is that a tweet from a customer on Twitter that talks about a service that the company needs to know about, for example, is free open text data that doesn't fit well into the tabular form of the RDBMS. Even if it did, says Olson, the questions the company will want to ask in a sentiment analysis of the natural language captured in the tweet, and any others that follow, simply can't be provided. "SQL doesn't have the words for that," notes Olson.

"You need a different kind of data platform that can look at very large amounts of information and do that and also can do different new kinds of processing.

The ongoing conversation right now is about whether to choose new technology aimed at problems rather than trying to shoehorn every single problem into the traditional RDBMS technology.

Olson argues for the new technology because there are problems that don't fit using RDBMS.

"As an old line data base guy what's interesting to me right now, is if you asked me in the 1980s or 1990s any question I would have answered relational database. That was the technology used for data management. These days the answer's different. That is, IT guys are buying Hadoop, they're looking at NoSQL products, you know, distributed hash tables other kind of storage engines; columnar stores is getting very big; the RDBMS conventional table stores they remain in a really important highly valuable market. But you're starting to see data centers choose technology that are aimed at problems rather than just the old mature technology," says Olson.

"If you are doing behavioral targeting you will want to pay attention to users' preferences, some relationship graphs, you will want to consider looking at sentiment that is embedded in user generated content on Web sites. You will want to bring to bear machine learning and pattern recognition and statistical analysis tools that have never made sense before, and were never available before. As the volume increases you can just lob another couple of servers into the cluster and scale up with it. It's nice, scalable, elastic, kind of the modern cloud computing.

"The data is driving a new kind analysis and I think that is really interesting in the enterprise today. We are not just collecting numbers and dates and character trends, we are collecting audio, video, sentiment and analyzing this stuff.

"The key thing is Hadoop is different in quality from those older products. So Hadoop is really good at exhaustive batch analysis deep processing data at scale. If you've got a data warehouse you're used to flying through your cubes interactively and very quickly getting answers to questions that are sort of recomputed. Those are two very important things. You want to do both. But they really are different. If you've got an existing enterprise data warehouse problem it is going to map poorly to Hadoop. Where Hadoop will really shine is as these new data sources come online – as you want to understand them, digest them, and maybe even load summaries about them into your cube Hadoop is the way you do that. So you don't want to say I've got a problem I've been solving with IBM DB2 for the last 20 years now I'm going to port it to Hadoop."

Accenture's Akred says that while most enterprise data management activities focused rightly on creating a single version of the truth around financial transactions and supply chains using RDBMS, the struggle to accommodate more unstructured data and semi-structured data forces enterprises to define the structure of that data when they want to introduce a user.

"So what Hadoop gives us is a combination of the ability to handle a wide array of different kinds of data – structured, unstructured, time series and the like – on one hand and then the ability to process its scale and deliver it to those enterprise systems so that we can look at that data in the context of the rest of the enterprise. So it's a combination of the ability to adjust and handle data at scale and then make that data available to the rest of the enterprise. Our processing and ETL and other processes like that that makes Hadoop a very valuable component in a larger enterprise data management solution."

BI typically uses data from transactional and other RDBMSs collected by the enterprise – such as sales and purchasing records, product development costs, and new employee hire records – cleans it for accuracy and consistency, then puts it into a form the BI system is programmed to run queries against, says a Price Waterhouse Cooper study. Such systems are necessary for accurately analyzing transactional information but they don't work well for messy questions coming from unstructured sources such as Twitter that haven't been able to scale to analyze large data sets efficiently.

"In contrast, big data techniques allow you to sift through data to look for patterns at a much lower cost and in much less time than traditional BI systems," says the PWC report. Big data approaches allow for more questions of more information, opening a wide range of potential insights enterprises couldn't afford to consider in the past.

"Furthermore, big data analysis is usually iterative: you ask one question or examine one data set, then think of more questions or decide to look at more data. That's different from the 'single source of truth' approach to standard BI and data warehousing," the PWC report says.

"Specifically, enterprises are also motivated by the inability to scale their existing approach for working on traditional analytics tasks, such as querying across terabytes of relational data. They are learning that the tools associated with Hadoop are uniquely positioned to explore data that has been sitting on the side, unanalyzed.

"The big data analysis supplements, not replaces, the BI systems, data warehouses, and database systems essential to financial reporting, sales management, production management, and compliance systems. The difference is that these information systems deal with the knowns that must meet high standards for rigor, accuracy, and compliance – while the emerging big data analytics tools help deal with the unknowns that could affect business strategy or its execution."

TDWI's Russom agrees that the Hadoop system (HDFS) has advantages over a DBMS in some circumstances. "As a file system, HDFS can handle any file or document type containing data that ranges from structured data (relational or hierarchical) to unstructured data (such as human language text)," says Russom. "When HDFS and MapReduce are combined, Hadoop easily parses and indexes the full range of data types. Furthermore, as a distributed system, HDFS scales well and has a certain amount of fault tolerance based on data replication, even

when deployed atop commodity hardware. For these reasons, HDFS and MapReduce (whether from the open source Apache Software Foundation or elsewhere) can complement existing BI/DW systems that focus on structured, relational data."

Furthermore, the MapReduce component of Hadoop brings advanced analytic processing to the data. This is the reverse of older practices where large quantities of transformed data were brought to an analytic tool, especially those based on data mining or statistical analysis. As big data gets bigger, it's just not practical (from both a time and cost perspective) to move and process that much data.

Gartner's Adrian says that while data warehouses have been the mainstay of big data analysis in the past, the question of whether they will remain so in individual enterprises depends largely on how often data is used.

A distributed file system on inexpensive hardware may still be an effective option for little used, rarely changed data now that tools such as MapReduce, an analysis component of Hadoop, are emerging to process it efficiently.

The HDFS is gaining popularity to store data on a massively parallel collection of inexpensive commodity hardware. Typically these approaches involve MapReduce, a programming model for processing and generating large data sets that is automatically paralleled and executed on a large cluster of commodity machines. The resulting programs are relatively specialized and may have very limited use. They may be tested and discarded when newer approaches come along, but they can be very cost-effective, says Adrian.

Use cases for MapReduce include extract, transform and load (ETL) processing, "sessionization" of Web logs and various types of data mining. In such cases, the result set may be imported into a data warehouse for further processing with SQL. Alternatively, MapReduce may be run inside databases. And finally, an emerging approach is the use of Hive, an SQL-like layer, atop Hadoop. Which approach to use, and when, has as much to do with an organization's skills and resources available as it does with the technology.

Although much of the big data conversation centers on entirely new workloads with data outside the warehouse, Adrian says, the fact is that data warehouses are handling problems of greater scale and complexity than many of the new use cases. He advises organizations to ask themselves what drives its best architecture? Start with where the needed data resides. If it's already inside a data warehouse, emerging techniques for processing inside, "close to the data," allow you to leverage the features of an RDBMS and the platform running it. New techniques using MapReduce – as well as user-defined functions and data types – are emerging and may let enterprise's innovate. Some products offer "sandboxing" to build temporary places to experiment inside your data warehouse and tear them down when finished.

Big data is driving innovative approaches like these that change as frequently as the problems do, says Adrian. He advises remaining open to new ideas, and frequent experimentation.

Clearly the challenges with big data acquisition and access are around the difficulty to manage large volumes of unstructured or semi-structured data in real-time or near real-time streaming velocities and multiple, often-dissimilar data sets. The fact that streaming velocities of real-time or near real-time data coming in to an organization require them to be addressed very quickly, requires a solution that can keep up and not fall behind due to low network bandwidth or slow compute performance.

The open source community has a number of software initiatives that have led exponents to boast of 2012 as "the" year of big data applications. Many of the innovations related to big data have come from Web 2.0 companies, yielding an ever-growing collection of open source technologies that dramatically scale out clusters of smaller inexpensive servers. These tools allow storage and analysis of data in ways that simply were not possible before. As interesting as these open source technologies and solutions have become, not all are suitable for the enterprise. While open source solutions are attractive from the standpoint of the innovation they can bring, many

organizations require dependable, supported, and tested enterprise-class solutions for rapid deployment and mission-critical operation.

Moreover, enterprises signing on to open source solutions should keep in mind the need to have enough specialized staff on hand to engage with them, potentially adding to their costs.

Considerations such as these favor solutions from big data industry leaders such as Cisco and Oracle who enjoy a long-term relationship based initially on creating secure networking solutions for Oracle environments that now expand to the overall data center and beyond. Including industry-specific solutions. With Oracle Solution Centers around the world using the Cisco Unified Computing System and Cisco Nexus data center switches, the two companies have tested and documented best practices for configuration and deployment of Oracle Solutions running on Cisco Unified Computing System including the new Oracle NoSQL Database, Oracle Relational Database, Real Application Clusters (RAC), Partitioning, Applications, WebLogic Middleware, Oracle VM and Oracle Linux.

Some enterprises, especially in the financial services industry, have been managing large data volumes for years. Existing data management products that have supported these enterprises are now adding new features to support big data requirements. Some IT professionals say platforms enterprises are using today are often adequate to manage the growth and variety of data. Storage optimization techniques such as data compression provide an effective solution to store and reuse petabytes of raw data. Organizations use data compression to use less space for the same amount of data, thus keeping operational costs from rising as the data accumulates.

The challenge is to be able to access data on demand without prejudicing performance. Some proven commercial products are available to manage higher data volumes, however, and maintain performance and reliability of the infrastructure.

Furthermore, new enhancements in large objects (LOB) columns allow enterprises to manage big data challenges while ensuring the security and reliability of their mission-critical systems.

Yet banks, despite large BI teams, are said to only be scratching the surface of big data. That will change as increasingly sophisticated high performance big data solutions come to market, the first being Oracle's NoSQL database technology on Cisco's UCS hardware. The fact that Oracle, the leading data management company in the world has released a highly engineered Apache Hadoop solution for global distribution on Cisco's UCS is a major validation of Apache Hadoop.

Winning the Gold with Big Data

A major reason for the growth of big data is financial. The idea that there's gold in an enterprise's data has spread through academic and analyst studies that suggest early adopters using big data to inform their decisions are more productive and win higher returns on equity than competitors that don't. Where content is king, content is money. Acknowledging this, early adopters have raised their investments to handle big data by 50 percent to $4 trillion since 2005, according to IDC.

The astronomical numbers relating to the size of the digital universe make a compelling case for organizations to accept that what they've been doing with their traditional relational databases and enterprise data warehouses (EDW) in order to remain competitive up to now just isn't going to cut it much longer. That is, if they're going to take advantage of the huge new benefits of big data, avoid the headaches, and turn it into real business value. This means shifting the perspective from data as structured and disciplined to extracting value from the vast, messy array of unstructured digital data engulfing us; a rethinking of the data deluge as a problem to the data deluge as an opportunity.

In 2011, the McKinsey Global Institute reported a study of big data in five market segments – retail and healthcare in the United States, the European public sector, and manufacturing and personal-location data globally. It concluded that big data can generate value in each. A retailer going full out with big data, for example, would add more than 60 percent to its operating margin. Similarly, the public sector would reap enormous benefit by creatively and effectively using big data to drive efficiency and quality. For the U.S. healthcare sector, the savings would be more than $300 billion in value every year.

Other sectors poised to gain substantially from using big data are the computer and electronic products and information sectors, as well as finance and insurance, and government.

McKinsey reminds us that all this digital information makes it possible to do many things that previously could not be done: spot business trends, create customer-centric organizations, manage risk and so on. Hernando De Soto, author of *The Mystery of Capitalism*, attributes the "brilliance of western capitalism" not to creating a formula for making money but to its "property memory systems". These are the result of examining, selecting and validating information about who owns land, labor, credit, capital and technology, how they are connected, and how they can be profitably recombined.

Making the most of big data means keeping the data in the first place, rather than discarding it.

"You may as well keep everything," says Cloudera's Olson. "Our contention is that all things being equal you don't know what is relevant, so it would be stupid to throw anything away. This is the lesson I think that we all learned from Google. There is so much out there it's beyond human imagination how much you could know how much information there is on the Web. What you do is basically, you search using smart tools to help you find the relevant data. So it's not about storing and discarding data any longer it's about finding and using the data on the Web. The platform we are building has to make that possible."

In some cases data is being simply thrown away for lack of an effective means of capturing, storing, and analyzing them. RDBMS with their comparatively rigid data models and transactional focus can be cumbersome for developers to adapt to the varying data types and flexible analysis models required of these data sets.

In fact, by some reckonings most organizations have ignored up to 95 percent of their available data because it has been too expensive to dip into until now. Big data makes it affordable to do so. To drive this point further a Forrester Research Inc. analyst defines big data as "extremely scalable analytics."

Nevertheless, many professionals consider managing unstructured data a major problem. They also believe that when done well and purposefully it can bring big rewards. But many people are still trying to manage their structured data, let alone their unstructured data, according to a survey conducted in mid-2011 by Joe McKendrick, an analyst at Unisphere Research, a division of Information Today, Inc. Eighty percent of survey respondents (446 data managers) in June, 2011 say big data is increasing in their organizations bringing clutter and challenges to their indexing and tagging.

"Many organizations are becoming overwhelmed with the volumes of unstructured information – audio, video, graphics, social media messages – that falls outside the purview of their 'traditional databases'," says McKendrick. Thirty-five percent of respondents say unstructured information has already surpassed or will surpass the volume of traditional relational data in the next 36 months. Sixty-two percent say this is inevitable within the next decade. The survey was gathered with input from readers of Database Trends and Applications magazine.

Research firm, IDC recognized the "consternation" among data center managers faced with capturing and analyzing big data, in a 2011 report. "Data center architectures and organizational models will need to evolve as big data applications pervade a company's infrastructure," says IDC in its report Extracting Value from Chaos, sponsored by EMC. The IT architectural and organizational approach used in newly emergent clustered

environments is radically different from the converged and virtualized IT environments driving most organizations' data center transformation strategies, says IDC.

"Big data will inject high-velocity requirements associated with capture and analysis, as well as results/predictive reporting. With big data, IT is best organized around the specific opportunity and/or capability rather than merely a set of shared services that serve both traditional and newer uses. Most IT disciplines – from infrastructure to applications to governance – are ideally part of a single integrated team and work closely with users of big data in ways that are very distinct from traditional enterprise IT approaches," say John Gantz and David Reinsel, authors of the IDC report.

Less than one in five organizations in the Unisphere Research survey that say they already have high levels of unstructured data think their current infrastructure will be sufficient to deal with big data effectively. Little more than a quarter will provide the same level or more to unstructured data. This may include services and capabilities such as storage, data security, monitoring, oversight, and data administrators.

Just under one-third of organizations surveyed could say with any certainty that the IT infrastructure and processes they maintain for relational data are also adequate for managing unstructured data.

Cost is one of the main barriers for 50 percent of respondents to effectively improving the management of unstructured data.

'Putting a Square Peg in a Round Hole'

Most companies with large concentrations of unstructured data are moving it into non-relational databases such as file systems, content management systems, or special purpose databases. But 42 percent deposit big data into relational databases, in many instances beside relational data. Says Unisphere's McKendrick, this shows the dysfunction and retention of the status quo, as people go to what they know – to the traditional RDBMS. "Unstructured data doesn't belong in relational structures. Putting it there is akin to trying to put a square peg in a round hole and will ultimately be expensive in the form of misappropriated and hard to find files," says McKendrick.

Yet 82 percent of the unstructured data-intensive organizations say they store big data in non-relational databases.

"This is a defective, sub-optimal process," adds McKendrick, "and may point to a perception on the part of some managers that storage and technology is the main challenge posed by unstructured data – versus viewing this data as presenting an opportunity to provide greater insights and resources to decision-makers, or a new business opportunity."

The largest group of respondents (37 percent) use search and access solutions against their content management systems to analyze or deliver unstructured data. One-quarter of the organizations surveyed use data mining and text analytics.

Among emerging solutions that better manage big data in respondents organizations are log monitoring and reporting tools (19 percent of respondents) in-memory databases (18 percent) NoSQL databases (17 percent). NoSQL databases are important for data storage, retrieval, and accommodating analytics for massive-scale data sets. These systems have been available for some time, but for big data analytics they have become particularly significant as cloud-based processing brings them to a level for smaller businesses, departments to use, and for occasional queries.

Cisco and Oracle: A Winning Combination

To address the needs of rapid acquisition and access of hundreds of terabytes of data to expose key insights and information hidden in vast volumes of data, Cisco and Oracle are the first companies to partner to deliver cost-effective, integrated, enterprise-class NoSQL database solutions that provide real-time big data acquisition and access with exceptional agility, availability, scalability, and manageability. Together these powerful and tightly

integrated components provide the building blocks to quickly deploy and scale effective big data infrastructure designed to provide high data availability.

Industry leaders, Cisco and Oracle have enjoyed a long-term relationship based initially on creating secure networking solutions for Oracle environments that now expands to the overall data center and beyond, including industry-specific solutions. With Oracle Solution Centers around the world using the Cisco Unified Computing System and Cisco Nexus data center switches, the two companies have tested and documented best practices for configuration and deployment of Oracle Solutions running on Cisco Unified Computing System including the new Oracle NoSQL Database, Oracle Relational Database, Real Application Clusters (RAC), Partitioning, Applications, WebLogic Middleware, Oracle VM and Oracle Linux.

Cisco is the worldwide leader in networking, transforming how people connect, communicate and collaborate. Together, Cisco and Oracle provide differentiated, scalable and available end-to-end solutions, while minimizing deployment risks, complexity, and total cost of ownership. Cisco and Oracle are uniquely positioned as global leaders in technology providing innovations for many shared customers. Oracle and Cisco locations around the world stand ready to support and help educate your staff on how to manage and leverage the unprecedented amounts of information, or big data, for competitive advantage. Learn more at www.cisco.com/go/oracle and www.cisco.com/go/bigdata.

Using the Oracle NoSQL Database and Cisco UCS, businesses get access to exceptional, scalable performance. This is backed up with impressive benchmark results for high throughput event processing.

Organizations needing to implement NoSQL at their own pace, while maintaining their existing RDBMSs and moving data between the two can work with both traditional Oracle Database running on Cisco UCS blade servers as well as Oracle NoSQL Database running on Cisco UCS Rack-Mount servers coexisting in a single (UCS) management domain, sharing data back and forth between the two databases in a seamless fashion. The data can then be moved to a batch-oriented system like Hadoop or a data warehouse RDBMS. The data could also flow the other way to provide a very fast key-value lookup front-end to an RDBMS.

Since many open-source NoSQL are in pre-production versions with key features yet to be implemented, the Cisco UCS and Oracle NoSQL database solution represents an enterprise-grade, integrated real-time big data solution. Cisco and Oracle integrate and optimize the software, compute, and network components with extensive testing and reference architectures that reduce risk. The combined hardware and software solution is fully supported by Cisco and Oracle, allowing organizations to deploy the Cisco UCS and Oracle NoSQL solution confidently and scale it predictably on proven compute and network infrastructure.

Since the challenge of adding a new node to a current big data (NoSQL) implementation over time exposes the customer to inconsistent set up leading to configuration and performance issues, Cisco UCS uses Cisco UCS Manager to manage the system and within this software is the ability to pre-define how a server is to be set up to run a specific workload. These configurations are called service profiles and these allow for not only consistent configurations but also allow for servers to be re-purposed in minutes to run different workloads. With Cisco UCS, the application/database workloads are really separated from specific server resources.

As a new tool, nearly every NoSQL developer is in learning mode. This learning curve can make it difficult to find experienced NoSQL programmers or administrators. Recognizing this, the comprehensive solution from Cisco and Oracle helps organizations deploy big data solutions quickly, with validated and supported configurations that scale easily and predictably as demand dictates.

Cisco's 10,000 Server Customers

"Cisco is on par – and many would argue they are superior – to where IBM and HP are with servers," says analyst Zeus Kerravala. "And it's not HP that Cisco has put a dent into with servers – it's IBM. The decision used to be between HP and IBM, and now it seems to be Cisco and HP," with Cisco's Unified Computing System platform.

Jennifer LeClaire, a prolific reporter-blogger reporting this, commented in January, 2012 that Cisco defied naysayers just two years after rolling out its UCS and won more than 10,000 data center managers and CIOs as customers.

"Two years ago, industry watchers doubted Cisco could gain traction with its new computing technology in a market dominated by Hewlett-Packard and IBM," LeClaire wrote. "But UCS is proving to be a force to be reckoned with, capturing 53 industry benchmark performance world records and winning a dozen industry awards for innovation since it started shipping in July, 2009."

"This announcement is a pretty big deal," said Kerravala, principal analyst at ZK Research. "I was really skeptical as to whether Cisco could be a server vendor at all. Two or three years ago, it was unusual for someone to deploy a Cisco server. It was like someone deploying Macs in the workplace. Now it's not that unusual. That 10,000 mark legitimizes Cisco as a server vendor... Cisco has almost taken IBM's place as an alternative to HP."

"Cisco seems to be in the right place at the right time – right in the middle of the cloud's growing adoption," LeClaire wrote. "The 2011 Cisco Cloud Index forecasts that cloud computing is transforming business, with more than 50 percent of computing workloads in data centers expected to be cloud-based by 2014.

"Cisco's Cloud Index also predicts global cloud traffic will grow 12-fold by 2015, to 1.6 zettabytes per year. That's the equivalent of more than four days of business-class video for every person on Earth.

"This explosive cloud growth requires advanced data center capabilities – and Cisco is promoting UCS to support end-to-end cloud application delivery."

Cisco's Good Timing

"UCS was designed from the ground up with what Cisco calls a 'clean slate' approach. It is an integrated system, designed to optimize computing, networking, storage access, virtualization and management. Cisco listened to CIOs' complaints and tapped into a market demand for a new approach to computing with UCS.

"Cisco would argue that it's the first server that's been optimized for virtualization as the way it handles memory, processors and Ethernet. The server was designed with cloud computing and virtualization in mind," Kerravala says. "A lot of the legacy servers out there were not."

UCS is the first fabric-computing platform that combines industry-standard, x86-architecture with networking and storage access into an integrated system. Cisco works with industry-leading infrastructure and software vendors, including BMC, CA, Citrix, EMC, Hitachi Data System, Microsoft, NetApp, Oracle, Red Hat, SAP and VMware to provide pretested, end-to-end solutions.

"Cisco's strength has always been in looking for market transitions that allow them to enter new markets. In this case the market transition is virtualization," Kerravala says. "Cisco's advantage is that they have designed a server specifically for this era of computing."

Cisco and Oracle's NoSQL Database Solution

Cisco UCS already delivers a proven, scalable, and flexible architecture for the Oracle RDBMS and Oracle applications, providing all-in-one infrastructure for Oracle Real Application Cluster (RAC) deployments. The compute nodes, cluster interconnects, and storage access all work together flawlessly and are managed from a single management domain. Partnerships with leading industry storage vendors such as EMC, NetApp, and others offer a choice of external enterprise-class storage arrays.

Unlike traditional systems, the database middle-tier system can also reside in the same management domain, offering rapid on-demand provisioning with click-of-the-mouse simplicity – significantly reducing IT budget.

The Cisco UCS and Oracle NoSQL Database solution seamlessly integrates infrastructure for running traditional Oracle Database on Cisco UCS with infrastructure for deploying Oracle NoSQL Database. For Oracle NoSQL Database, the solution provides powerful Cisco UCS C-Series Rack-Mount Servers, connected with Cisco UCS 6200 Series Fabric Interconnects. This support enables existing Oracle RDBMS configurations running on Cisco UCS to be combined with an Oracle NoSQL Database deployment, controlling everything under a single Cisco UCS management domain.

Less than two years after being introduced, Cisco UCS has become a mainstream application platform with a large number of customers, a long list of world-record benchmarks, and support for major platforms including Oracle, Microsoft, Red Hat, and SAP. Cisco UCS is the first unified system available anywhere, combining industry-standard x86-architecture servers with networking and storage access into a single management domain that incorporates both blade and rack-mount servers.

Asked to give an example of a Cisco product that came from breaking down its customers' silos, John Chambers, Cisco CEO and chairman answered blade servers. Chambers went on to tell Forbes' Rich Karlgaard in January, 2012, "it's the only computing technology that can handle data, voice and video. We looked into blades seven years ago, talked to our customers and fully committed four years ago. Now we have the number two market share and growth at 100 percent. At the time people said, 'Cisco in blade servers?' "

Cisco has a major role to play in the big data universe, and not just in the network. Cisco is one of the few vendors that can provide a complete infrastructure stack, end-to-end, to build big data environments and integrate them seamlessly into the rest of the enterprise. Cisco's network infrastructure has been powering some of the largest big data clusters in the world to date, and the company is extending that experience into the compute side of the equation with its UCS servers. It's forging partnerships with leaders in the big data arena – for example, its UCS servers and Nexus switches were recently certified for Cloudera's Hadoop and Oracle's NoSQL database. Its knowledge of the network gives it a unique perspective into big data clusters which, although they attempt to minimize network traffic by computing data locally, still place significant demands on the network during ingest, data shuffle, and replication, for example.

Still, there's resistance. "It's extraordinarily hard for people to change from making decisions based on personal experience to making them from (big) data – especially when that data counters the prevailing common wisdom," warns the Winter 2011 MIT Sloan Management Review. The solution is to think biggest, the Review advises.

As the volume and interconnectedness of data vastly increases, the value of the big data approach will only grow.

If the amount and variety of today's information is daunting, the world in five or 10 years will be much more so. The deluge of data from Facebook, Twitter, location-based services, and other forms of social media, is only one part of it. Elsewhere, the stakes are much higher: While advertisers and consumers are focused on monetizing sites that have hundreds of millions of users for a few pennies each, the ubiquity of connectivity and the growth of sensors has opened up a larger storehouse of information that will not only help businesses profit, but will also boost safety and enable environmental benefits.

Take a typical passenger jet that generates 10 terabytes of information per engine every 30 minutes of flight. In a single six-hour, flight from New York to Los Angeles on a twin-engine Boeing 737, the total amount of data generated is a massive 240 terabytes of data. With the total number of commercial flights approaching 30,000 in the U.S. on any given day, a day's worth of sensor data quickly climbs into the petabyte scale. Multiply that by weeks, months, and years, and the number is colossal.

"The volume of data we're generating now from machines pales in comparison to the volume of data we'll soon generate from our own bodies," says data security expert Dave Asprey. Individuals are becoming mobile sensors – collecting, creating, and transmitting information, from locations to body status to environmental information. This is happening as smartphones equipped with cameras, microphones, geolocation, and compasses proliferate. Wearable medical sensors, small temperature tags for use on packages, and other radio-equipped sensors are a reality. Insight-oriented analytics from this ocean of information – where interactions cause tides and waves in the flow and delivery of business value – will become a critical competitive requirement. Big data technology is the likeliest path to gaining such insights.

Acknowledgements: This document was written by Chris Forsyth with contributions from Ramesh Chitor, Business Development Manager, Cisco Global Enterprise Partner Organization.

Chris Forsyth
chris4syth@gmail.com