Big Data: Big Opportunities to Create Business Value

Report and recommendations based on discussions with the Leadership Council for Information Advantage

Council Members
- **Rich Adduci**, Vice President and Chief Information Officer, Boston Scientific
- **Dave Blue**, Senior Manager, Enterprise Data Services, The Boeing Company
- **Guy Chiarello**, CIO, JPMorgan Chase
- **John Chickering**, Vice President, Fidelity Investments
- **Dimitris Mavroyiannis**, Group Chief Information Officer, Eurobank EFG Group
- **Sanjay Mirchandani**, Senior Vice President and Chief Information Officer, EMC Corporation
- **Joe Solimando**, Senior Vice President, Global Operations and Technology, CIO, Disney Consumer Products
- **Deirdre Woods**, Associate Dean and CIO, The Wharton School, University of Pennsylvania

Special Contributors
- **Johann Schleier-Smith**, Co-Founder and Chief Technology Officer, Tagged.com
- **Ian Willson**, Boeing Technical Fellow and IT Enterprise Architecture, The Boeing Company

An industry initiative sponsored by the Information Intelligence Group of EMC
About the Leadership Council for Information Advantage

The explosion of mobile networks, cloud computing and new technologies has given rise to incomprehensibly large worlds of information. The rapidly shifting dynamics of competition, coupled with a deluge of data, create new challenges for leaders across all sectors who want to tap into the power of information to make better and more timely decisions about how their companies can best compete, grow and create new sources of value.

Information has become a potent source of competitive advantage—on par perhaps with an organization’s capital assets and human talent. This is why EMC has convened the Leadership Council for Information Advantage, an advisory group made up of global information leaders from “information-advantaged” enterprises, organizations from a variety of industries that are successfully using information to revolutionize how they compete and do business. We conduct periodic, in-depth interviews with Council members and publish their ideas in a series of reports that provide candid lessons learned, proven best practices, and expert guidance on how to transform information into business value.

This report, the third in a series from the Leadership Council for Information Advantage, provides executive-level guidance on how organizations can begin enacting the organizational changes required to derive value from “big data”—techniques and technologies that make handling data at extreme scale efficient and affordable. Future reports will explore other topics and strategies for leveraging information to gain business advantage.

Information advantage defined

Information advantage is about cultivating the mindset, skills, processes and technologies to use information to operate more efficiently, increase customer loyalty, grow market share and create business opportunities that were not possible before.
# Contents

**Big Data: Big Opportunities to Create Business Value**

Executive Summary ........................................................................................................................................... 3  
What is big data? ........................................................................................................................................... 5  
Our Digital World: New Data Sets, New Possibilities .................................................................................. 5  
Enhancing BI with big data: achieving “high-def” business visibility ........................................................... 6  
Rethinking Data Wisdom: When More is More ............................................................................................ 7  
For Tagged.com, big data is the heart of product innovation ........................................................................... 7  
Building Infrastructures for Big Data ........................................................................................................... 8  
Big data in the cloud ................................................................................................................................... 10  
Big Data Road Map: Strategies for Success ................................................................................................... 11  
Mindshifts and Jobshifts: The Democratization of Information ....................................................................... 14  
Conclusion .................................................................................................................................................. 15  
Leadership Council for Information Advantage Member Biographies ....................................................... 17
Executive Summary

The speed at which business moves today, combined with the sheer volume of data created by the digitized world, requires new approaches to deriving value from data. Hidden inside streams of structured and unstructured data are answers to questions that businesses haven’t even thought to ask or haven’t been able to ask because of technological limitations. Because of the speed of business today and the massive amounts of data being generated, organizations must find new ways of getting to data—figuring out what’s in it and what to do with it. Recent advances in storage, network and compute technologies enable organizations to economically and efficiently harness big data and turn it into a potent source of business advantage.

Forrester Research estimates that organizations effectively utilize less than 5 percent of their available data. This is because the rest is simply too expensive to deal with. Big data technologies and techniques represent an important advancement, because they make it efficient and affordable for organizations to tap into the 95 percent of data currently passing them by. Imagine the upside: if two companies use data with the same effectiveness, but one handles 15 percent of available data while the other is stuck at 5 percent, which enterprise is more likely to win?

When used correctly, big data can yield insights to develop, refine or redirect business initiatives; discover operational roadblocks; streamline supply chains; better understand customers; as well as develop new products, services and business models.

Some examples of big data in use:

• The U.S. federal government collects more than 370,000 raw and geospatial datasets from 172 agencies and subagencies. It leverages that data to provide a portal to 230 citizen-developed apps, with the aim of increasing public access to information not deemed private or classified.

• Professional social network LinkedIn uses data from its more than 100 million users to build new social products based on users’ own definitions of their skill sets.

• Silver Spring Networks deploys smart, two-way power grids for its utility customers that utilize digital technology to deliver more reliable energy to consumers from multiple sources and allow homeowners to send information back to utilities to help manage energy use and maximize efficiency.

• Jeffrey Brenner and the Camden Coalition mapped a city’s crime trends to identify problems with its healthcare system, revealing services that were both medically ineffective and expensive.

While the usefulness of big data may be clear, the path toward big data productivity is not. Successfully leveraging big data insight requires a real investment in proven technologies, updated workforce skills and leadership focus. Organizations must combine three facets of strategy—technical, organizational, cultural—in order to implement a big data platform that suits the business and its objectives.

The Leadership Council on Information Advantage outlines a roadmap for companies to move up the big data learning curve and better leverage their information:

1. Start with what you already have. By preparing existing data stores for analysis through integration, tagging, and other methods, the overwhelming task of collecting new data should take a back seat to working with existing data sets initially. Organizations will need a well-considered strategy for integrating big data into their information architecture so that it becomes a core part of the platform and the way they do business.
Council members recommend organizations assemble a team of business and technical leaders focused on big data to think through these issues and plan for the opportunity.

2. Line of business leaders and IT professionals should work together on identifying which existing data pools have the greatest value. Pick and choose areas of business in which new insight would offer the most impact, prioritize the related data for analysis and build test cases. Find people who are passionate about the business and get them invested early on.

3. Once a few test cases have produced results, begin exploring different uses and combinations of data to create new insight. Spark the imagination of business leaders to ask previously unexplored questions and the creativity of the IT department to overlay disparate data types in new ways.

4. To ensure the insight gained is actionable, address any potential security, privacy, compliance or liability issues early on. Consider how big data techniques differ from traditional ones, and review and update data policies accordingly. Be sure that all concerns regarding the source, use and results of data manipulation have been addressed. Also, organizations need to think creatively about revamping business processes and work flows to take advantage of what’s learned from big data.

5. Cultivating the human capital to take advantage of big data opportunities and insights can be more challenging than cultivating the right big data technologies and processes. Organizations will need to augment their talent rosters with data scientists—people who combine business acumen with analytical creativity and technical expertise. Big data specialists will be required to bridge business and IT, and their skill sets will have to extend well beyond traditional DBMS and BI.

At the end of the day, it’s what organizations do with their big data insights that make a difference. Capitalizing on big data will require profound changes in the way organizations view the role of data within the enterprise. Directors should reorganize departments to promote data-driven decision making, ensuring the instruments for capturing data are in place and encouraging unrestricted manipulation of data to unveil insight. IT organizations must accommodate storing and working with big data, and make available analysis tools that are approachable, easy to work with and integrated into business processes.

Big data will make a big difference in the coming years. Senior executives should begin considering how their companies can benefit from new insight derived from big data.

What is big data?

Big data is not a precise term; rather it’s a characterization of the never-ending accumulation of all kinds of data, most of it unstructured. It describes data sets that are growing exponentially and that are too large, too raw or too unstructured for analysis using relational database techniques. Whether terabytes or petabytes, the precise amount is less the issue than where the data ends up and how it is used.

“My belief is that data is a terrible thing to waste. Information is valuable. In running our business, we want to make sure that we’re not leaving value on the table—value that can create better experiences for customers or better financial results for the company.”
—Johann Schleier-Smith, Tagged.com
Our Digital World: New Data Sets, New Possibilities

There's no end in sight for the proliferation of data. With enterprise data volumes moving past terabytes to tens of petabytes and more, business and IT leaders face unique opportunities to capitalize on this data for competitive advantage. Companies that align their processes, operations and corporate culture to embrace and exploit big data will gain the benefit of timely, differentiated insight; those that do not risk falling by the wayside.

According to IDC’s 2011 Digital Universe Study commissioned by EMC, the amount of information created and replicated this year will surpass 1.8 zettabytes (1.8 trillion gigabytes), growing by a factor of nine in just five years. It's interesting to note that the amount of information created by individuals themselves—documents, photos, music files, blog posts, etc.—is far less than the amount of information being created about them in the digital universe, according to the study. Data about data, or metadata, is growing twice as fast as the digital universe as a whole.

Web sites alone generate staggering amounts of data. Facebook has more than 800 million active users, and there are more than 900 million objects (pages, groups, events and community pages) that people interact with. Facebook users spend over 700 billion minutes per month on the site, creating on average 90 pieces of content and sharing 30 billion pieces of content each month. Facebook’s data infrastructure team is responsible for quickly analyzing all of that data to present it to users in the most relevant way, and to understand preferences, uses and sentiment as a basis for launching new products.

As Facebook demonstrates, big data enables innovative business models, products and services. It gives companies a way to outperform the competition. According to a May 2011 McKinsey Global Institute report, a retailer embracing big data has the potential to increase its operating margin by more than 60 percent.

Companies are leveraging these and many other sources of data to achieve a better understanding of their customers, employees, partners and operations, with an eye towards improving every aspect of business. In fact, the Leadership Council for Information Advantage anticipates that data will generate a similar productivity boost to the enterprise that IT has over the past 20 years. Big data has the potential to redefine business, and companies that understand this stand to become leaders in the global marketplace.

“Data growth is a factor that everybody is trying to deal with. We're seeing tremendous growth in the size of data, year on year, as I think everyone is. Finding effective approaches for containing cost so that it doesn't run away with your budget is an issue. Another major challenge is dealing with unstructured data. How do you manage that data effectively? How do you control its growth? And how do you actually make that data part of the information fabric that people can draw upon to make decisions or look at information?”

—Rich Aducci, Boston Scientific
Enhancing BI with big data: achieving “high-def” business visibility

The term “big data” encompasses more than structured and transaction-based data. It also includes videos, RFID logs, social networking conversations, sensor networks, search indexes, environmental conditions, medical scans, “data exhaust”—the trail of clicks through the Internet produced by web surfers—and more. Anything that can be digitized can produce data about who is using it, how they are using it and possibly even why they are using it. Big data isn’t always new data; sometimes it’s existing data looked at in different ways. Today there is more data being produced than computer networks are capable of transporting.

Big data techniques complement business intelligence (BI) tools to unlock value from enterprise information. Whereas BI traditionally performs structured analysis and provides a rear-view mirror into business performance, big data analytics provides a forward-looking view, enabling organizations to anticipate and execute on opportunities of the future.

Simple reporting, spreadsheets and even fairly sophisticated drill-down analyses have become commonplace expectations of BI. However, there are types of analyses that BI can’t handle, particularly when data sets become increasingly diverse, more granular, real-time and iterative, requiring organizations to capture in-depth information from a specific moment in time before conditions change rapidly. These types of unstructured, high volume, fast-changing data—big data—breaks the relational database model. Such data requires a new class of technologies and analytic methods to extract value. For example, big data approaches are essential when organizations want to engage in predictive analysis, natural language processing, image analysis or advanced statistical techniques such as discrete choice modeling and mathematical optimization—or even if they want to mash up unstructured content and analyze it with their BI mix.

Companies that augment BI with big data stand to gain a more holistic view of business. It’s like going from analog television with only the basic network channels to high-definition TV with premium cable. The result for organizations is “high definition” visibility into business conditions that yields rich, wide-ranging, more accurate and actionable insight that can help address customer needs, operational risks and performance opportunities, both within the enterprise and the extended supply chain. With big data analysis, companies can gain understanding not just about what’s happening with the business and why, but to also comprehend what else is possible.
Rethinking Data Wisdom: When More is More

The old way of thinking says too much data is a bad thing, as accumulating data drives up infrastructure costs and becomes unruly to manage and mine. Today, companies are realizing that more is better, as big data offers new ways of making money, driving efficiency, and gaining a competitive edge.

The benefits of big data are beginning to be realized by companies in every sector of the global economy. However, some industries will have an easier time leveraging big data than others. Those sectors that tend to invest aggressively in IT will be better equipped to handle the technology shift. And industries accustomed to relying on data to inform the business will adapt more quickly to big data and will be more practiced at converting that data into insight. Council members anticipate early adopters of big data analytics will include the financial services, retail, manufacturing and Internet media industries.

Ultimately, the benefits of big data are timely, in-depth business insights. It will take time and new ways of thinking to get there, so business leaders should begin considering the infrastructure, personnel, and culture changes their organizations must undergo.

Building Infrastructures for Big Data

With the entirely new rates of data growth happening, everything must be re-evaluated. Enterprises are making significant investments in new infrastructure to capture, store, aggregate, manage, govern and analyze data—an undertaking that must be approached holistically and with big data analytics in mind. To accommodate the data itself, IT infrastructures must be able to inexpensively store higher volumes and more types of data than ever before. Data velocity, meaning the speed at which data changes, must also be accommodated.

Such massive quantities of data are difficult to move swiftly back and forth over today’s network connections. Big data infrastructures must distribute compute power so that data analysis can be done close to the user to avoid the latency inherent in crossing networks. Distributing this kind of compute power to fuel analysis tools and provide real-time responses will pose challenges as organizations realize that analysis may well need to occur at the place where the data resides. Council members see a trend in which the velocity and volume of data make it impractical to move data back and forth for processing. Instead, computation and analytics will likely move to the data. Furthermore, Council members see cloud computing models becoming essential to the success of big data projects (see sidebar on page 10).
For Tagged.com, big data is the heart of product innovation

Tagged.com takes a fresh approach to social networking—instead of connecting people with other people they know, Tagged introduces them to people they might like to know through a portfolio of products including a dating service, games, photo sharing and chat. A startup without the legacy baggage of a traditional enterprise, the company leverages big data to make associations and has been profitable since 2008.

One example of a Tagged product powered by big data is Meet Me, the network’s dating service that provides a photo and a brief description of two people and asks if they want to meet. If they both agree, they’ve made a match. The system makes the determination of which of Tagged’s 100 million user profiles should be presented to each user.

“One of the things that we’ve found to be really helpful in this regard is to study that graph of interactions, who’s friends with whom, who talks to whom and so forth,” says Johann Schleier-Smith, the co-founder and CTO of Tagged.com, who says Meet Me has been a major contributor to the more than tenfold growth of the company. “We’re able to make it really personalized to the individual, for each user at a time, to make a recommendation for who you might want to connect with.”

Tagged collects 50 billion log entries every month on five billion page views, corresponding to about ten terabytes of data.

Tagged uses this data to make predictions—will people who meet form a match?—and to run ad-hoc analyses to understand customer behavior—at what point will a gamer pay to play? These things fall outside the scope of traditional business intelligence (BI) tools. Because the data the social network generates is clickstream data in a homogenous format, the company can easily move data between systems to compile creative mash-ups and perform analytics.

“We really have to have this ability to ask arbitrary questions, which is provided by the database, and then have analysts and smart people who are product-oriented, who are thinking about customers, who are thinking about the business and who are asking questions,” Schleier-Smith explains. “The key thing for us is that they’re able to get answers to those questions and they’re able to get those answers very quickly … then they go on to ask the next question. This is really key to a lot of our business decision making when we’re deciding what types of (products to launch), what should our next game be, or why is this one working? We really want to understand what the customers are doing, what matters to the customers.”
But storing and serving up the data is not enough—it must be synthesized, analyzed, and correlated in new ways in order to deliver business value. Some big data techniques require working with data that hasn’t been modeled by data architects, allowing for the comparison of different types of data and pattern matching across disparate data sources. This allows big data analytics to provide new perspectives on traditional corporate data and yield insights into data that traditionally has not been analyzed. Tools such as Hadoop, open-source technology that distributes data-analysis workloads across many computers to break analysis into many parallel workloads and produce results faster, are essential. Commercial tools are still nascent, as big data is still a fairly new phenomenon. As a result, most of the software programs used today for big data analytics are purpose-built and developed in-house using open source tools created by the Apache Software Foundation, Google, Yahoo and others.

An early example is Walmart, which last year launched Walmart Labs to track social media outlets and gather data about the retailing giant’s products and brands. Scale is important when dealing with social media feeds, and no off-the-shelf software could meet Walmart’s needs. So, the company built a tool loosely based on Google’s MapReduce for tracking web pages. Called MapUpdate, Walmart’s tool lets the company rapidly map a huge amount of data to track things like what are Twitter users’ interests on a particular day, and use that information to influence features such as predicting customers’ future purchases. The company believes that gleaning near real-time information from social media outlets and incorporating that into future purchase predictions—instead of basing predictions on purchase histories that don’t always indicate future interests—will give the company an edge over its online competition.

When allowed to influence business, big data drives results. Progressive Casualty Insurance Company recently introduced its Snapshot program that offers discounts to safe drivers not based on past driving records, but on current driving behaviors. Customers of the program enroll online and install a plug-in device in their car that tracks the number of miles driven, the time of day customers typically drive and how often they brake hard. Based on that information, Progressive can offer discounts for safe drivers of up to 30 percent. By using basic driving data in an innovative way, the company attracts new customers who pride themselves on safe driving with an incentive to try out Progressive’s auto insurance policies.

Big data in the cloud

Cloud models tame big data while extracting business value from it. This delivery model gives organizations a flexible option for achieving the efficiencies, scalability, data portability and affordability they need for big data analytics.

Cloud models encourage access to data and provide an elastic pool of resources to handle massive scale, solving the problem of how to store huge data volumes and how to amass the computing resources required to manipulate it. In the cloud, data is provisioned and spread across multiple sites, allowing it to sit closer to the users who require it, speeding response times and boosting productivity. And, because cloud makes IT resources more efficient and IT teams more productive, enterprise resources are freed to be allocated elsewhere.

Cloud services specifically designed for big data analysis are starting to emerge, providing platforms and tools designed to perform analytics quickly and efficiently. Companies who recognize the importance of big data but don’t have the resources to build the required infrastructure or acquire the necessary tools to exploit it will benefit from considering these cloud services.

“Real-time data will continue to grow at a faster pace than the capability to move it. Unless we change the way we address the problem, we are going to find ourselves constantly struggling to squeeze information through very narrow tubes. I believe that we’re going to be faced with a situation where more and more we have to do the analytics where the data resides. Instead of moving the data for processing, we are going to move analytics closer to the data.”

—Dimitris Mavroyiannis, Eurobank EFG Group

“The cloud will play an important role in big data. I think it’s going to be increasingly rare that you’re going to be able to run all this [infrastructure] at home. And why would you, in some cases?”

—Deirdre Woods, The Wharton School of the University of Pennsylvania
Big Data Road Map: Strategies for Success

Strategies for dealing with big data challenges will differ depending on the “data maturity” of the organization. How efficiently and effectively can data be collected for analysis, and for that matter, is the organization aware of all of the different types and sources of data that should be included to maximize insight and answers? How well can the organization reconcile different data formats? What is the cost of collecting and analyzing data, and how is that cost weighed against the anticipated value of the outcome?

Business leaders are beginning to face the issue of how to take advantage of big data in a way that’s appropriate for them. Learning how to amass and manage large data sets and work with them to glean value in a way that’s realistic and achievable requires strategic planning and forethought.

Members of the Leadership Council for Information Advantage advise approaching big data by leaning into the data-learning curve. This means working with existing data stores to prepare them for a new type of analysis, but also working among departments to set data priorities, goals and limitations. It also means evaluating the level of data competency in the organization—both within IT and in lines of business—and determining where skills can be transferred or enhanced to leverage big data.

“We’re working in an era where data has got to be accessible when people want to exploit it. To analyze some big data sets, it means making a big gamble on infrastructure—building it out, getting all the data in one place, and then exposing it to the relevant set of users—takes time and resources. By the time you get it all done, you might miss the whole window, especially with Internet-based data, which changes at light speed. You can use cloud-based services to rapidly get some applications in place that can provide access to your data. Then, you can test things out and make smaller bets (on big data).”
—Joe Solimando, Disney Consumer Products

“Obviously, the cost of storage technology is dropping, so we, as an industry, have been able to allow data to proliferate and we’ve been able to store it. But then you look at the enterprise in the marketplace being faced with competition, and it’s clear the winner will either be the one that is lower cost or higher effectiveness, or some appropriate point on the value curve between the two. Those who can operate more efficiently or more effectively are the ones who are going to succeed. So content and process management helps with making sure you have the right data available at the right time, and it isn’t slowing you down or driving up your operating costs. And so content management, done right, gives the enterprise an edge over the competitor who had all the same big data, but just isn’t doing the content management right.”
—John Chickering, Fidelity Investments

“I think it’s true what they say about 90 percent of analysts’ time being spent gathering the data instead of doing the analysis that they’re skilled at ... and figuring out how to act on it.”
—Ian Willson, The Boeing Company
Council members recommend:

1. **Control data hygiene.**
   Starting with the information already stored in corporate IT architectures, organizations should clean up existing data stores to prepare them for this new form of analysis. Basic block and tackle “data hygiene” such as compressing, de-duping and archiving old files will streamline storage, enable outdated systems to be retired and make it easier to identify data stores that need to be brought up to date. Also, integrating data wherever possible, implementing a data tagging system, and training IT staff in partitioning data are important parts of the preparation process.

   Businesses will benefit from exploring how to mitigate big data storage growth and contain complexity and cost. For example, compiling the data required to complete a report on a certain product or service could mean storing the data components on different tiers—the most expensive data in a traditional data warehouse, the data of lesser value on commodity storage, and other data in analysis tools—and pulling those components together on the fly.

2. **Pinpoint data value.**
   Bring together line of business leaders and IT practitioners to identify which data pools have the greatest value. Evaluate the data stores that have been prepared for analysis and consider how they could be expanded or improved, and look at unstructured data sets and prioritize which ones should be converted to more usable formats. Business leaders and IT must also work together to highlight use cases for the data based on existing business to determine which approaches will yield the most business value in the shortest window.

   Look for areas in business where growth is desired and begin asking questions of big data. Instead of focusing on results, focus on getting answers to those questions and working with that information. Once answers to one situation emerge, expose them to business users who are likely to come up with new uses for the data and new scenarios to pose, and the usefulness of the data grows organically. As insight grows, IT is inspired to expand the data sets, cross referencing them in new and different ways. And so the cycle builds.

   It’s also advantageous to involve people who are passionate about the business and eager to explore new options when starting out. People who are invested in

   “Because it has become so cheap and easy to store data, a lot of companies have operated under this idea of, ‘let me just store it, I’ll deal with it when I figure out how to deal with it.’ But now the velocity of growth is increasing. The amount of storage we’re using is proliferating. All of that compels us to bring a business discipline to this ecosystem that helps us understand what needs to be retained and for how long. Big data raises the stakes on why content management needs to be promptly and successfully woven into the business operation. Then it’s as simple as basic records management blocking and tackling.”
   —John Chickering, Fidelity Investments

   “With compute power being what it is … you don’t need to build big tables and land them on disks and keep them on disks. You build them on the fly. That has reduced data storage needs dramatically. It’s a form of, I guess, intellectual compression, not algorithmic compression. It’s just smart data modeling and using the power of what you’ve got.”
   —Ian Willson, The Boeing Company

   “Big data calls for a lot more creativity in how you use data. You have to be way more creative about where you look for business value: if I combined this data with this data, what could it tell me?”
   —Joe Solimando, Disney Consumer Products

   “Instead of waiting for big data to stop operations, we should better organize or archive our data, manage it over its lifecycle and actually get rid of it. You can move out of mitigation mode by doing a better job of managing your information up front—in other words putting the data to more efficient use.”
   —David Blue, The Boeing Company
succeeding will be inspired to ask creative and thoughtful questions of big data, and the results will reflect that. Successful test cases of big data analytics will develop a level of comfort and confidence that can be leveraged going forward.

There are, of course, economic considerations. In a perfect world free from budgets, every piece of data that is collectable would be collected, and every byte would be analyzed in as many ways as the mind can consider. But in reality, collecting, storing, and analyzing data comes at a cost. Companies will need to make economic decisions about which data is worth collecting and analyzing. And different parts of the business will have to make compromises. Business leaders are likely to lean towards collecting and analyzing more data, while IT leaders well aware of technology budget limitations and staff restrictions may lean in the other direction. Given the iterative nature of big data, these decisions will need to be revisited on a regular basis to ensure the organization is considering the right data to produce insight at any given point in time.

The more data collected, the bigger the economic problem becomes. It’s more expensive to store and manipulate more data, and the more data there is to process the more computational power is required, layering on more cost. Yet more data produces better informed decisions. Approaching big data analytics with finite definitions of what data will be considered sounds counterproductive, but companies—especially those just starting out on big data projects—will need to set some parameters around just which data is involved and gauge expectations of results accordingly.

3. Dimensionalize your data mix.
As businesses progress along this data learning curve, they can begin exploring new uses and combinations of data. This means collecting new types of data, adding new sources of data to existing sets and combining sets to create new value and insights.

For example, Coca-Cola’s Freestyle next-generation beverage dispenser that serves 125 different flavors of drinks sends information such as which brands are most popular during what time of day back to company data specialists for analysis. Being able to gather this usage data from various locations and combine it with existing inventory information allows Coca-Cola to better stock even its non-Freestyle dispensers with the right amount of product at the right time of day.

Business workers should be encouraged to use their imagination to test hypotheses and validate or disprove hunches with the help of big data. IT should also get creative in pioneering new ways to collect, partition, and combine data so that insight is unveiled and action can be taken.
Address data portability, security, privacy, compliance and liability issues up front, so that the insight generated can be acted on without consequence. For companies in highly regulated industries, adherence with regulations also needs to be re-examined. Council members cited variations in data privacy laws among different countries as a factor to consider in working with data sets from multiple jurisdictions.

Because the technologies for handling big data are still evolving, Council members point out that standards-based approaches will become especially important to avoid problems with sharing data between systems or transferring it from one platform to another. Companies must also explore how combining data in new ways might lead to new security or privacy concerns. When integrating third-party data into analytics projects, what are the reuse risks and liability concerns? If you build commercial products using third-party data, could you be held liable for problems resulting from errors in such data?

5. Prepare to act on what’s learned.
All information-driven insight—regardless of whether it’s from a team of management consultants or from big data analytics—is only as valuable as what is done with it. Big data offers organizations the opportunity to derive detailed, timely insights and act on them with greater speed and agility than ever before. For instance, analyzing social media data could uncover customer behaviors to customize promotions and offers presented to certain customers. Achieving this type of real-time responsiveness to opportunity will require organizations to become far nimbler about how they manage business processes and workflows. Business leaders must set expectations of action and charge managers with injecting data-driven discoveries into how their teams work. Council members concede that creating the organizational flexibility to adapt may be the toughest challenge of all.

“Every vendor wants to bundle everything and provide a one-stop shop. This makes sense, and I don’t have a problem with that, as long the vendor doesn’t lock me in into a specific solution. And I think the only way that this can be avoided is if the vendor follows industry standards. We’ve moved into a world where standards prevail, especially in big data analytics, where data originate from multiple sources. Vendors that provide standard-based big data solutions are much more likely to be preferred.”
—Dimitris Mavroyiannis, Eurobank EFG Group
Mindshifts and Jobshifts: The Democratization of Information

Council members believe most organizations will find the hardest part of adopting big data analytics is not the technology itself, but cultivating the human capital to take advantage of it. Almost every Council member cited difficulty in finding data analysts, data engineers or data scientists with both the technical acumen and business insight to drive big data projects. Data analysts may also require additional training to adjust to a new world of analytics with big data, although confidence is high that many BI professionals possess portable skills and will make use of emerging tools that make it easier to analyze and work with big data.

Even more challenging than the shortage of skilled data analysts is cultivating the collective imagination of an organization to leverage big data for business insight. Several Council members predicted this obstacle would prove to be far more enduring and intractable than any technology considerations.

Prepare for a mindset shift, not just a technology shift, advises Council members. Unlike previous trends, the adoption of big data will likely be felt by many departments in an organization, not just IT. Before, analysis meant scouring small data sets and making formal queries on cleansed data to find an answer. In the future, data warriors in all departments of an organization will focus on data from mixed sources to improve decision making.

Council members foresee a day in which big data tools will be deployed to business users across the organization, empowering them to self-provision data sets and conduct queries without IT intervention. Propelling the shift to technical self-sufficiency is the consumerization of IT. Many business users today are technically astute and quite comfortable using new tools. IT departments will be able to train business workers on analytical tools so that reports, dashboards, and other instruments of information can be updated by the workers themselves, leaving IT to focus on more strategic elements of technology.

By doing this, IT departments empower business workers to create their own knowledge. When analysis happens at various levels in the organization companies promote self-service solution finding. And by allowing queries to be generated by business workers who are closer to the data in the first place, a whole new range of question possibilities and points of view generate richer, more contextual solutions.

Working with business users will expand the capabilities of IT workers, bringing them closer to the strategic goal of aligning business and IT. Business workers will gain a better understanding of the capabilities and limitations of technology.

“At EMC we are developing roles for what we call the data scientist: people with a good amount of data competence who have skill sets partitioning information to make it easier to work with. The capabilities people in this role bring in the value chain of an organization are pretty tremendous. The central function is to add core business value and mask (for business users) the heavy lifting that happens behind the scenes in IT.”
—Sanjay Mirchandani, EMC

“The biggest challenge is the people. Big data requires non-traditional IT skill sets. We’re bringing in more Ph.D.s and people with expertise in outside areas to help our business users work with information.”
—Guy Chiarello, JPMorgan Chase
Conclusion

Big data is a disruptive force that will affect organizations across industries, sectors and economies. Not only will enterprise IT architectures need to change to accommodate it, but almost every department within a company will undergo adjustments to allow big data to inform and reveal. Data analysis will change, becoming part of a business process instead of a distinct function performed only by trained specialists. Big data productivity will come as a result of giving users across the organization the power to work with diverse data sets through self-service tools.

And that’s just the beginning. Once companies begin leveraging big data for insight, the action they take based on that insight has the potential to revamp business as it is known today. If a marketing department can gain immediate feedback on a new branding campaign by analyzing blog comments and social-networking conversations, do focus groups and customer surveys become obsolete? Nimble new companies that understand the value of big data will not only challenge existing competitors, but may also begin defining the way business is done in their industries. Customer relationships will undergo transformation as companies strive to quickly understand concepts that previously couldn’t be captured, such as sentiment and brand perception.

Achieving the vast potential of big data calls for a thoughtful, holistic approach to data management, analysis and information intelligence. Across industries, organizations that get ahead of big data will create new operational efficiencies, new revenue streams, differentiated competitive advantage and entirely new business models. Business leaders should begin thinking strategically about how to prepare their organizations for big data—and big opportunities.

“A few years ago, I would’ve said the value in BI overshadowed that of big data, but now I’d say the relationship is even, if not reversed. There’s more outside information now that can be digitized—about user behaviors and external conditions—that can be layered on top of structured data. This type of analysis opens a window into not just what happened and why, but it also helps you see what’s possible.”

—Guy Chiarello, JPMorgan Chase
Biographies for the Leadership Council for Information Advantage and Special Contributors

**Rich Aducci**  
*Vice President and Chief Information Officer*  
*Boston Scientific*

Rich Adduci joined Boston Scientific in 2006 as CIO, where he is focused on integrating multiple IS organizations into a single global team of IS professionals focused on enabling competitive advantage through innovative use of information and technology. He also serves as a member of Boston Scientific's operating committee, quality management board and capital committee. Prior to joining Boston Scientific, Mr. Adduci was a partner at Accenture, where he led the company's health and life science practice. He holds more than 15 European patents and two U.S. patents for the development of modeling tools to support business strategy and market entry for new wireless technologies. Mr. Adduci holds a BS in industrial engineering from Purdue University and an MBA from the University of Chicago with concentrations in finance and economics.

**Dave Blue**  
*Senior Manager, Enterprise Data Services*  
*The Boeing Company*

Dave Blue leads the delivery of enterprise data services within Boeing Information Technology. He previously led Boeing Information Architecture, where he was responsible for developing and communicating the vision, strategy, and architecture supporting information management disciplines and for applying this architecture to projects. As a member of the Chief Architects Council (CAC), he helped ensure that information architecture was integrated within the overall enterprise architecture. Mr. Blue's Boeing career has been in information technology, with progressively broader responsibilities in application development and maintenance, information management, and architecture disciplines.

**Guy Chiarello**  
*CIO*  
*JPMorgan Chase*

Guy Chiarello has worldwide responsibility for information technology at JPMorgan Chase. He joined the firm in 2007 and is a member of its Executive Committee. Previously, Mr. Chiarello was Morgan Stanley's Chief Technology Officer and Chief Information Officer for seven years, responsible for strategy and execution for the global IT organization. He served in numerous other IT roles during his 23 years at Morgan Stanley, including two years working for the Office of the Chairman. Mr. Chiarello began his IT career in 1981 with the Treasury Department for the State of New Jersey. For more than a decade, Mr. Chiarello has been an executive advisor for leading public technology companies on business strategy and technology innovation and remains very committed in this area. He is also very active in the emerging technology landscape, influencing innovation roadmaps and investments throughout the venture community. The enhanced focus on innovation that Mr. Chiarello brings to JPMorgan Chase has helped the firm garner numerous technology awards, including the 2010 Chair’s Choice award for Innovation in Custody & Securities Services Technology, Profit & Loss 2010 Digital FX Awards for Best Interest Rate Platform and Best Corporate Platform, as well as many awards for Chase mobile payment and banking solutions, including the firm's iPhone and Android mobile banking apps, the Quick Deposit feature and Instant Action Alerts. Mr. Chiarello has garnered industry and private sector recognition through various awards, including Top Financial IT Executive by CIO Forum, Computerworld Premier 100 Leaders, CIO of the Year by NASSCOM and Information Week Top Innovators. He is the Vice Chair on the Board of NPW, a technology advisor to PENCIL, an Executive Board leader of the Leukemia and Lymphoma Society of Central New Jersey and an active fund raiser for the Cancer Institute of New Jersey. Mr. Chiarello is a graduate of The College of New Jersey with a B.S. in business. He was recognized with The Distinguished Alumni Citation Award and has recently been distinguished with a special Citation for Academic and Athletic Excellence.
John Chickering  
Vice President  
Fidelity Investments  
John Chickering’s experience as a consultant, software vendor, end user and lecturer gives him a unique perspective on applying technology to manage information. He is a vice president at Fidelity Investments, where he currently works on electronic delivery of customer communications. Mr. Chickering has delivered solutions in the public sector and financial services industries and has served as CIO at two human resource services companies. A former licensed merchant marine engineering officer, he began his IT career at American Management Systems, where he was a founding member of the firm’s imaging practice. After nearly 10 years, he moved on to spend two years at a workflow software vendor before joining Fidelity. Mr. Chickering has authored several articles and has spoken at both industry conferences and continuing education seminars hosted in academia. He is a member of AIIM’s Board of Directors, where he is serving as the Board’s Chair for 2012. He is also an active volunteer with several community service organizations. Mr. Chickering holds an MBA (Operations Research) from the University of Maryland and a BS (Marine Engineering) from the United States Merchant Marine Academy.

Dimitris Mavroyiannis  
General Manager—Group CIO  
Eurobank EFG Group  
Dimitris Mavroyiannis oversees all of Eurobank EFG Group’s IT units, ensuring the various units are working as a whole to help achieve the bank’s overall business objectives, as well as maximize the value of IT investments, optimize the utilization of IT resources, and assure information systems and the technological infrastructure are able to support the company’s innovative business initiatives. Mr. Mavroyiannis joined the bank in 1999 to develop its Internet strategy and banking channel, a role that evolved into his leading a subsidiary specializing in e-business and e-commerce consulting and implementation services for the Greek market. Mr. Mavroyiannis was the CEO of this services group until 2004. He then served in various leadership roles for Eurobank EFG Group, including CIO of the bank’s operations in Greece. Prior to Eurobank EFG Group, Mr. Mavroyiannis worked for IBM Consulting Group in Europe, as well as for smaller companies in Greece and abroad. He has an MBA from Imperial College London, an MSc from University College London, and a BEng from the University of Sussex.

Sanjay Mirchandani  
Senior Vice President and Chief Information Officer  
EMC Corporation  
Sanjay Mirchandani is responsible for extending EMC’s operational excellence and for driving technological innovations to meet the current and future needs of EMC’s business. He also leads EMC’s network of global delivery centers in India, China, Russia, and Israel. These centers support EMC’s worldwide research and development efforts and provide customer support and shared services. Mr. Mirchandani most recently served as the senior vice president leading the EMC Office of Globalization. In this role, he identified global growth opportunities and built the EMC processes and infrastructure required for global expansion. He was also responsible for bringing in new strategic international partners into EMC’s Global Alliances program. Prior to joining EMC, Mr. Mirchandani was Microsoft’s Regional Vice President, Enterprise Services, Asia, where he worked with the region’s largest customers and partners. He also held multiple management positions during his tenure with Microsoft, including President, Asia Pacific Region; President, South Asia; and Managing Director, India. Mirchandani earned an MBA from the University of Pittsburgh and a BA from Drew University.
Johann Schleier-Smith
Co-Founder and Chief Technology Officer
Tagged.com

Johann Schleier-Smith is co-founder and CTO at Tagged. He is responsible for building up and expanding the social network that enables and inspires anyone to meet and socialize with new people. He has developed products used by millions, created and managed large-scale infrastructures, innovated software development techniques and built recommendation engines and machine learning systems that move the industry forward. Starting out as an entrepreneur in college, Mr. Schleier-Smith launched a dozen businesses in collaboration with co-founder Greg Tseng before focusing on social networking in 2004. He pursued a Ph.D. in Physics at Stanford for several years and holds an A.B. in Physics and Mathematics from Harvard University.

Joe Solimando
Senior Vice President, Global Operations and Technology, CIO
Disney Consumer Products

Joe Solimando defines the strategic direction for information technology across all of Disney Consumer Products' (DCP) lines of business, which include licensing for toys, apparel, and hardlines products; retail stores; worldwide publishing; and e-commerce. He also serves as DCP’s segment representative on The Walt Disney Company IT Leadership Board, which oversees The Walt Disney Company’s information technology direction, standards, and company-wide IT initiatives. Mr. Solimando joined Disney in 1998 as vice president, operations & technology of Disney Consumer Products. In this role, he managed DCP’s Shared Applications Services group responsible for the implementation and support of shared financial and HR business applications. He also led the planning, development, and implementation of operations and technology systems for several business units as the IT business partner for vertical businesses, including Walt Disney Art Classics, Disney Direct Marketing, Walt Disney Records, and Disney Worldwide Publishing. Prior to joining Disney, Mr. Solimando held the position of senior manager of information technology in the management consulting practice at Ernst & Young. In his 10 years with this firm, he worked on IT strategic planning, system evaluation, selection and implementation projects for many top consumer products, retail, entertainment and manufacturing companies. Mr. Solimando has also held information technology and project management positions at Wicke’s Companies and Fluor Engineers. He holds both an MBA and a BS in Civil Engineering degree from The Pennsylvania State University.

Ian Willson
Technical Fellow for Data Warehousing & Business Intelligence
The Boeing Company

Ian Willson is a former researcher and aviation software entrepreneur, who created the first consumer air travel software and Travel$ense, the industry standard for business travel analytics. His current focus at Boeing is the Common Data Warehouse being developed for the new 787 Dreamliner aircraft. Mr. Willson leads the database and technical architecture teams to integrate 50 internal and external authoring systems to create an integrated repository for all aspects of Boeing’s new aircraft programs, from conception through delivery and support. Previously, Mr. Willson designed Boeing’s first active data warehouse, improved its reporting efficiency 9,700 percent.

Deirdre Woods
Associate Dean and CIO
The Wharton School, University of Pennsylvania

Deirdre Woods leads a 120-person organization at Wharton Computing in developing and maintaining technologies that further The Wharton School’s leadership in research, knowledge creation, and teaching. In her years at Wharton, Woods has been instrumental in bringing student and faculty satisfaction with IT services to the highest level and has served as a strategic driver for some of Wharton Computing’s most innovative technologies. As Associate Dean and Chief Information Officer, Woods ensures that all of the school’s various technology initiatives and programs are effectively implemented