Big Data is the Future of Healthcare

With big data poised to change the healthcare ecosystem, organizations need to devote time and resources to understanding this phenomenon and realizing the envisioned benefits.

Executive Summary

Big data is already changing the way business decisions are made – and it's still early in the game. However, because big data exceeds the capacity and capabilities of conventional storage, reporting and analytics systems, it demands new problem-solving approaches. With the convergence of powerful computing, advanced database technologies, wireless data, mobility and social networking, it is now possible to bring together and process big data in many profitable ways.

Big data solutions attempt to cost-effectively solve the challenges of large and fast-growing data volumes and realize its potential analytical value. For instance, trend analytics allow you to figure out what happened, while root cause and predictive analytics enable understanding of why it happened and what is likely to happen in the future. Meanwhile, opportunity and innovative analytics can be applied to identifying opportunities and improving the future.

All healthcare constituents – members, payers, providers, groups, researchers, governments, etc. – will be impacted by big data, which can predict how these players are likely to behave, encourage desirable behavior and minimize less desirable behavior. These applications of big data can be tested, refined and optimized quickly and inexpensively and will radically change healthcare delivery and research. Leveraging big data will certainly be part of the solution to controlling spiraling healthcare costs.

Simply by witnessing how big data has transformed consumer IT, it is clear that the promise of big data in healthcare is immense (think Google, Facebook and Apple's Siri, which all rely on processing and transmitting massive amounts of data). While its potential in healthcare has not been fulfilled, the question is not if, but when.

This white paper will define big data, explore the opportunities and challenges it poses for healthcare, and recommend solutions and technologies that will help the healthcare industry take full advantage of this burgeoning trend.

What Is Big Data?

A large amount of data becomes "big data" when it meets three criteria: volume, variety and velocity (see Figure 1). Here is a look at all three:
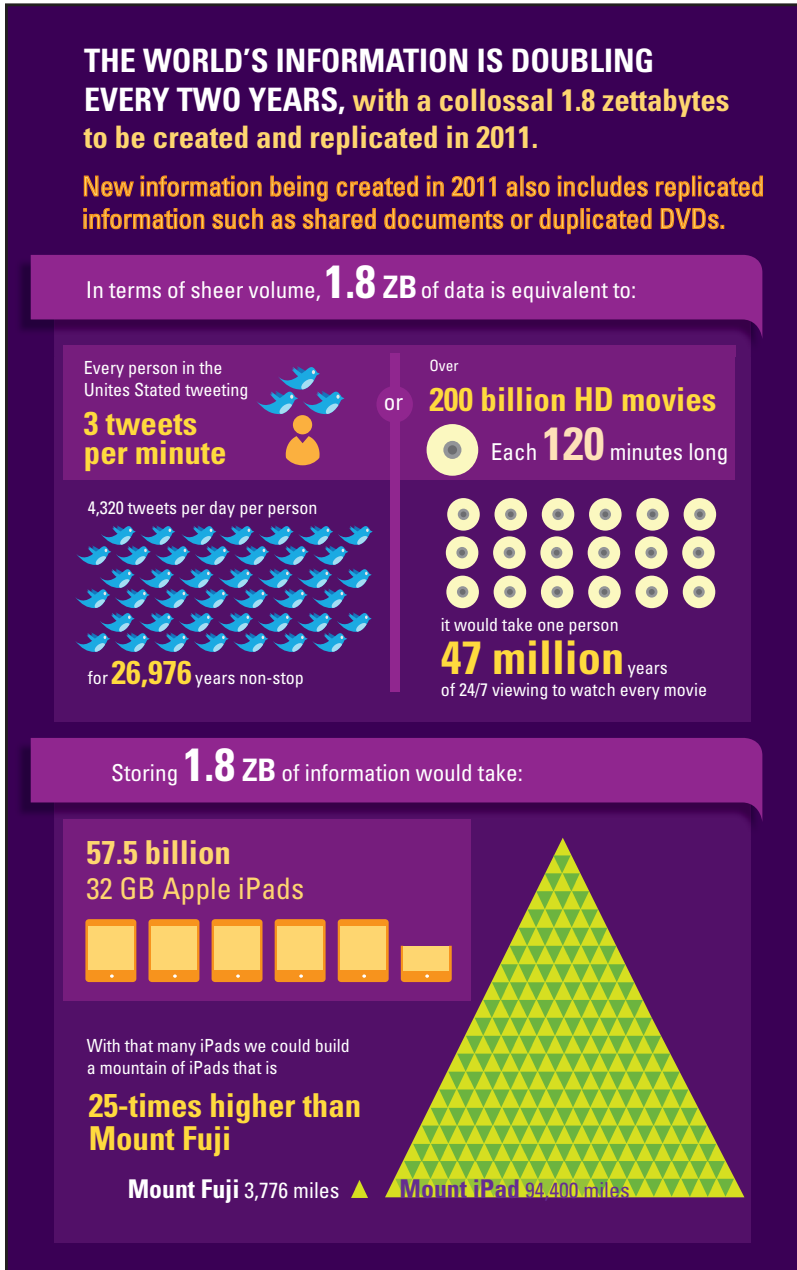
- **Volume:** Big data means there is a lot of data – terabytes or even petabytes (1,000 terabytes). This is perhaps the most immediate challenge of big data, as it requires scalable storage and support for complex, distributed queries across multiple data sources. While many organiza-

Cognizant

tions already have the basic capacity to store large volumes of data, the challenge is being able to identify, locate, analyze and aggregate specific pieces of data in a vast, partially structured data set.

- **Variety:** Big data is an aggregation of many types of data, both structured and unstructured, including multimedia, social media, blogs, Web server logs, financial transactions, GPS and RFID tracking information, audio/video streams and Web content. While standard techniques and technologies exist to deal with large volumes of structured data, it becomes a significant challenge to analyze and process a large amount of highly variable data and turn it into actionable information. But this is also where the potential of big data potential lays, as effective analytics allow you to make better decisions and realize opportunities that would not otherwise exist.

What Big Data Looks Like

THE WORLD'S INFORMATION IS DOUBLING EVERY TWO YEARS, with a collossal 1.8 zettabytes to be created and replicated in 2011.

New information being created in 2011 also includes replicated information such as shared documents or duplicated DVDs.

In terms of sheer volume, **1.8** ZB of data is equivalent to:

Every person in the Unites Stated tweeting

3 tweets per minute

4,320 tweets per day per person

for **26,976** years non-stop

or

Over

200 billion HD movies

Each **120** minutes long

it would take one person

47 million years of 24/7 viewing to watch every movie

Storing **1.8** ZB of information would take:

57.5 billion
32 GB Apple iPads

With that many iPads we could build a mountain of iPads that is

25-times higher than Mount Fuji

Mount Fuji 3,776 miles ▲ ▲ **Mount iPad** 94,400 miles

Source: "Extracting Value from Chaos," IDC Universe study, 2011; Mashable.com,
http://mashable.com/2011/06/28/data-infographic/
Figure 1

- **Velocity:** While traditional data warehouse analytics tend to be based on periodic – daily, weekly or monthly – loads and updates of data, big data is processed and analyzed in real- or near-real-time. This is important in healthcare for areas such as clinical decision support, where access to up-to-date information is vital for correct and timely decision-making and elimination of errors. Current data is needed to support automated decision-making; after all, you can't use five-minute-old data to cross a busy street. Without current data, automated decisions cannot be trusted, forcing expensive and time-consuming manual reviews of each decision.

Big Data = Big Opportunities

Big data has many implications for patients, providers, researchers, payers and other healthcare constituents. It will impact how these players engage with the healthcare ecosystem, especially when external data, regionalization, globalization, mobility and social networking are involved (see Figure 2).

Bringing the Patient into the Loop

The healthcare model is undergoing an inversion. In the old model, facilities and other providers were incented to keep patients in treatment – that is, more inpatient days translated to more revenue. The trend with new models, including accountable care organizations (ACO), is to incent and compensate providers to keep patients healthy.

At the same time, patients are increasingly demanding information about their healthcare options so that they understand their choices and can participate in decisions about their care. Patients are also an important element in keeping healthcare costs down and improving outcomes. Providing patients with accurate and up-to-date information and guidance rather than just data will help them make better decisions and better adhere to treatment programs.

In addition to data that is readily available, such as demographics and medical history, another data source is information that patients

Sectors Positioned for Greater Gains from Big Data

Historical productivity growth in the U.S., 2000-2008.
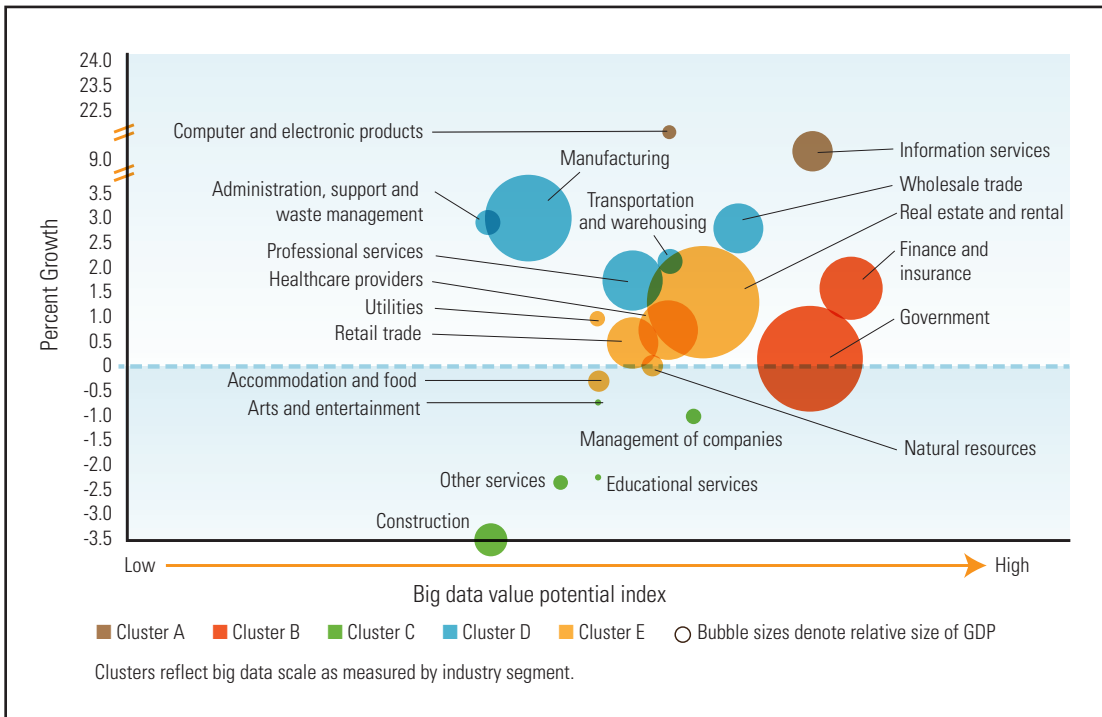
Source: U.S. Bureau of Labor Statistics; McKinsey Global Institute
Figure 2

divulge about themselves. When combined with outcomes, high-quality data provided by patients can become a valuable source of information for researchers and others looking to reduce costs, boost outcomes and improve treatment. Several challenges exist with self-reported data:

- **Accuracy:** People tend to understate their weight and the degree to which they engage in negative behaviors such as smoking; meanwhile, they tend to overstate positive behaviors, such as exercise. These inaccuracies can be accounted for by adjusting these biases and – through big data processing – improve accuracy time.

- **Privacy concerns:** People are generally reluctant to divulge information about themselves because of privacy and other concerns. Creative ways need to be found to encourage and incent them to do so without adversely impacting data quality. Effective mechanisms and assurances must be put into place to ensure the privacy of the data that patients submit, including de-identification prior to external access.

- **Consistency:** Standards need to be defined and implemented to promote consistency in self-reported data across the healthcare system to eliminate local discrepancies and increase the usefulness of data. Usage guidelines follow standards.

- **Facility:** Mechanisms based on e-health and m-health – such as mobility and social networking – need to be creatively employed to ease members' ability to self-report. Providing access to some de-identified data can simultaneously improve levels of self-reporting as a community develops among members.

Improving Quality with External Data

As progress is made toward initiatives such as electronic health records (EHR), more and more external data will become available, and this will become an integration challenge. External sources include the National Health Information Network (NHIN), health information exchanges (HIE), health information organizations (HIO) and regional health information organizations (RHIO). As sources and volume of information increase, so will expectations.

In addition to integrating data within the healthcare system, there are many potential benefits of integrating data from outside of the healthcare system. While integrating external data poses similar challenges to integrating internal data, there are also additional challenges, such as privacy, security and legal concerns, as well as questions about authenticity, accuracy and consistency.

As an example, external data about healthy people holds immense potential value for research and the future delivery of healthcare. Typical healthcare data includes only people visiting doctors and hospitals, which biases that data toward people seeking treatment. Adding anonymous data from large numbers of healthy people could help establish baselines, draw correlations and help with understanding the nature of illnesses. More data, effectively used, leads to better information and decisions, and more meaningful efforts.

Implications of Regionalization, Globalization

External data will come from different medical systems in various regions and countries. Effectively working across these disparate data repositories can help identify local knowledge and best practices and leverage them regionally and globally. Aggregating data regionally and globally also provides healthcare researchers with larger populations for clinical studies, trending and disease monitoring for epidemics, as well as early detection and the potential for improved results.

As data becomes less local and more regional and global, the quality of both data and metadata will improve over time as a result of increased data scrutiny and the efforts and contributions of big data innovators across the broader healthcare data ecosystem. At the same time, sharing data on a global basis will lead to security challenges, as well as issues resulting from different standards, terminology and language barriers.

Information Demands Drive Mobility

In many domains, mobility is a solution looking for a problem. Big data changes that. Demand for ubiquitous access to information mandates mobility and other technologies that provide access on demand. As data becomes more current, it will be necessary to get information into the hands of people with an immediate need for it, such as for clinical decision support. Users will also demand access to this data so they have precise and complete information to make the best possible healthcare decisions. Quality of care and improved outcomes will be the ultimate benefits.

Big Data, Social Media and Healthcare

Social media will increase communication between patients, providers and communities – e.g., patients with similar conditions and providers with similar specialties. This will not only work to globalize and democratize healthcare, but it is also a potentially important source of big data. Social networking data poses challenges such as volume, lack of structure and velocity, as well as new challenges around integration and accuracy.

For example, if a group of patients is discussing quality of care about a provider, there will likely never be 100% consensus. Patient experiences will be different, and there will be biases based on accidents, misunderstandings and other factors. The challenge will be to create useful information out of this collection of data to provide information such as provider ratings and improvement guidance.

Big Data = Big Challenges

The problem in healthcare isn't the lack of data but the lack of information that can be used to support decision-making, planning and strategy. As an example, a single patient stay generates thousands of data elements, including diagnoses, procedures, medications, medical supplies, lab results and billing. These need to be validated, processed and integrated into a large data source to enable meaningful analysis. Multiply this by all the patient stays across the system and combine it with the large number of points where data is generated and stored, and the scope of the big data challenge begins to emerge. And this is only a small part of the healthcare data landscape.

Outlined below are some of the specific challenges of healthcare big data, including healthcare as a technology laggard, data fragmentation, security, standards and timeliness.

Healthcare as a Technology Laggard

Healthcare is notoriously slow to redefine and redesign processes and tends to be a laggard in adopting technology that impacts the healthcare system, outside of some specific areas such as care delivery and research. In addition, the healthcare technology landscape includes vast areas of legacy technology, causing further complications.

Fragmentation

In healthcare, big data challenges are compounded by the fragmentation and dispersion of data among the various stakeholders, including payers, providers, labs, ancillary vendors, data vendors, standards organizations, financial institutions and regulatory agencies. Solutions for big data will break the traditional model, in which all data is loaded into a warehouse. Data federation will emerge as a solution in which the big data architecture is based on a collection of nodes within and outside the enterprise and accessed through a layer that integrates the data and analytics.

The biggest obstacle to effective use of big data is the nature of healthcare information. Payers, providers, research centers and other constituents all have their own silos of data. These are fundamentally difficult to integrate because of concerns about privacy and propriety, the complex and fragmented nature of the data, as well as the different schemas and standards underlying the data and lack of metadata within each silo. Even if everyone shared their data, there would be enough challenges integrating it within the silo, much less outside it.

Although groups such as HIE, RHIO and NHIN are working to facilitate the exchange of healthcare data, adoption has been slow, as they have faced numerous challenges.

Security

The entire healthcare system can realize benefits from democratizing big data access; for example, researchers can more easily collaborate, engage in peer review and eliminate duplication of efforts. Researchers will also be able to more readily identify opportunities where they can contribute and collaborate.

The cloud makes exposing and sharing big data easy and relatively inexpensive. However, significant security and privacy concerns exist, including the Health Insurance Portability and Accountability Act (HIPAA). A credentialing process could facilitate and automate this access, but there are complexities and challenges. Since providers, patients and other interested parties such as researchers need secure access, data access should be controlled by group, role and function. Finally, the security of the data once it leaves the cloud also needs to be assured. Big data can be used to identify patterns and irregularities indicating and preventing security threats, as well as other types of fraud.

Standards

Dealing with the myriad of standards (and lack thereof) creates interoperability challenges, at least through the medium term. Big data solution

architectures have to be flexible enough to cope with not only the additional sources but also the evolution of schemas and structures used for transporting and storing data. To ensure analytics are meaningful, accurate and suitable, metadata and semantic layers are needed that accurately define the data and provide business context and guidance, including appropriate and inappropriate uses of the data. This evolution of standards will eventually improve data quality.

Timeliness

Data timeliness is a challenge in various healthcare settings, such as clinical decision support, whether for making decisions or providing information that guides decisions. Big data can make decision support simpler, faster and ultimately more accurate because decisions are based on higher volumes of data that are more current and relevant. In some cases, there is a very limited window for clinical decision support – significantly smaller than the time it takes to run a report or analytic query. Careful attention to data and query structure, scope and execution is needed to ensure that the constraints of the processing windows are observed while still obtaining the best possible answer.

In other cases, streams of data containing complex and varied events without an overarching structure need to be mined. In this case, those events have to be turned into meaningful measures in real time that are, in turn, suitable for rapid analysis. In many cases, the only practical solution is to discard most of the data after analyzing it and selectively store those results. It's a tradeoff between the competitive advantage gained from the shorter feedback loop and the quality of the information that is being fed back. Capturing only processed data, streaming or otherwise, results in a loss of data at the expense of creating information. The underlying principle of big data is to keep everything, but in some cases that's just not practical or even useful – sometimes the hoarder reflex has to be checked and rational decisions made.

Big Data = Technology Choices

There are numerous technology solutions for dealing with big data, ranging from on-site to cloud and from open source to proprietary. On-site options that can tame big data include

Teradata, Vertica (HP) and Netezza (IBM). All of these solutions tend to have low time to value and maintenance but relatively high total cost of ownership.

Cloud-hosted software as a service (SaaS) solutions can help reduce the barriers of participating in the big data arena. Google and Amazon implement MapReduce-based solutions to process huge datasets using a large number of computers – e.g., terabytes of data on thousands of computers. MapReduce algorithms take large problems and divide them into a set of discrete tasks that can then be distributed to a large number of computers for processing and the results combined into a problem solution. Other cloud-based solutions include Tableau, which supports visualization.

Open-source Hadoop is a framework used by many companies as a high-performance, scalable and relatively low-cost option for dealing with big data. Training, professional services and support are needed to effectively deploy Hadoop solutions using the open source framework. Vendors such as Greenplum (a division of EMC), Microsoft, IBM and Oracle have commercialized Hadoop and aligned and integrated it with the rest of their database and analytic offerings.

SaaS is an important technology for democratizing the results of big data. SaaS-based solutions allow healthcare entities that control subsets of data to expose access through services that eliminate some of the aggregation and integration challenges. Additional services that facilitate analytics, both basic and advanced, can be made part of the overall offering.

Recommendations

To successfully identify and implement big data solutions and benefit from the value that big data can bring, healthcare organizations need to devote time and resources to visioning and planning. This will provide the foundation needed for strong execution. Without this preparation, organizations will not realize the envisioned benefits of big data and will risk being left behind competitors.

Our recommendations for healthcare organizations looking to leverage big data include:

- Establish a business intelligence center of excellence with a focus on big data.

- Decide on an appropriate big data strategy based on the organization's current and target business and technological maturity and objectives.

- Assess the various big data initiatives that can be deployed to meet overall corporate objectives, focusing initially on quick wins.

- Work with a partner that understands the full range of big data technologies and implications, including trends, security, internal and external system integration, hosting and development platforms, and application and solution development.

About the Author

Bill Hamilton is a Principal with Cognizant Business Consulting's Healthcare Practice, with nearly 20 years of experience in management and IT consulting across various industries. Bill has extensive experience in health plan strategy, operations and program management in the areas of transformation, modernization, information management and regulatory compliance. Technical areas of expertise include enterprise system design and development, mobile and distributed computing, database and data warehouse design and development, service-oriented architecture, cloud computing and software development lifecycle management and governance. Bill has written seven books and published articles about software development and database technologies. He can be reached at William.Hamilton@cognizant.com.

About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process outsourcing services, dedicated to helping the world's leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 50 delivery centers worldwide and approximately 145,200 employees as of June 30, 2012, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world. Visit us online at www.cognizant.com or follow us on Twitter: Cognizant.

Cognizant

World Headquarters
500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277
Email: inquiry@cognizant.com

European Headquarters
1 Kingdom Street
Paddington Central
London W2 6BD
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7121 0102
Email: infouk@cognizant.com

India Operations Headquarters
#5/535, Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060
Email: inquiryindia@cognizant.com