# ENTERPRISE IMAGING



Big data has the potential to define the future of medicine, guiding clinicians along the way in the very delivery of value-based care.

Dr. Shrestha is Vice President, Medical Information Technology, University of Pittsburgh Medical Center, Pittsburgh, PA; and Medical Director, Interoperability & Imaging Informatics, Pittsburgh, PA.

Disclosures: Dr. Shrestha is a Founding Member Executive Advisory Program at GE Healthcare, is on the Medical Advisory Board of Nuance, Inc., and Vital Images, Inc., as well as on the Editorial Board of Applied Radiology, and the Advisory Board of KLAS Research.

# Big data and cloud computing

Rasu B. Shrestha, MD, MBA

If you can dream it, you can do it.

- Walt Disney

ig data is more than just a buzzword in today's business vernacular - it is an opportunity in the clinical arena to bring previously unfathomable amounts of data to life, such that we can transform the data to valuable insights. It is an opportunity to put the data to work for us. The technologies around big data and the accessibility to these technologies made available via emerging cloud-based methodologies have the potential to transform healthcare in ways we have not even begun to dream about. Big data technologies will illustrate novel ways to measure improvements in quality care, patient outcomes, and will drive efficiencies in clinical workflow with new insights that we did not know were possible to attain. Big data, increasingly ubiquitous as a service in the cloud, indeed has the potential to define the future of medicine, guiding clinicians along the way in the very delivery of value-based care.1

## So, what is big data?

With the steady move away from analog, paper and film to digital, paperless and filmless, we have seen how even in healthcare, we today have a tremendous tsunami of data in multiple systems and databases across multiple hospitals, IT departments, and increasingly remote server farms often referred to as the "cloud." This digital content continues to grow exponentially. The University of Pittsburgh Medical Center (UPMC) currently has over 5.4 petabytes worth of data, and this is growing at a rate that will double every 18 months. In addition to traditional data sources, we are also seeing the rapid emergence of truly large stores of data coming in from content rapidly being created in social media (such as Twitter feeds), the world wide web, and medical and biological data being spewed out from genomic sequencers (such as Next Generation Sequencers, NGS). Increasingly, too, we are seeing more data from a spectacular array of medical devices, fitness devices, sensors, and interconnected equipment. This web of devices and equipment churning out huge amounts of data is also being referred to as the "Internet of Things" (IoT). So big data in many ways is a by-product of "big science"! The term "big data" refers to datasets whose size is well beyond the ability of traditional database software tools to capture, store, manage, and analyze.<sup>2</sup> The term refers not just to the data itself but also to an emerging set of newer technologies being developed to handle massive collections of data stores.

## Terabytes, petabytes, and exabytes

The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s.<sup>3</sup>

Big data, by definition, is an evolving measure in the sense that there is not one measure that quantifies big data, which could range from a few dozen terabytes (thousands of gigabytes) to many petabytes (thousands of terabytes) of data in a single data set. There is an assumption that as technology advances, the size of data sets that qualify as big data will increase. This amount can vary across industries, and can go into exabytes of data. If you are curious, one exabyte is one quintillion bytes, or one thousand petabytes, or quite simply one billion gigabytes. Since the dawn of time, till the early 2000's, it has been estimated that mankind created 2 exabytes of data. Across the globe, we create upwards of 2.5 exabytes of data every day,<sup>3</sup> and this is growing as human beings and machines go about their normal routines - communicating, browsing, documenting, capturing content, sharing, searching, and, well, living.

As one can imagine, the challenges with data this big can be many, from capture to curation, to management and processing of the data within a tolerable timeframe. However, we are fortunate to be living at a time when the technologies to process and analyze these massive stores of data are upon us. The opportunity at hand then is to cost effectively manage and analyze all of this data, and to connect the dots between disparate data elements to derive meaningful insights with the freedom, flexibility, and capability that one could previously only dream of.

### Variety is the spice of life: Big data's many V's

It is interesting to see the definition of big data grow "bigger." What was originally the 3 Vs of big data, now has at least 2, if not more, additional Vs latched on to it. Gartner's original definition of big data focused on the first three of these Vs. Per Gartner,<sup>4</sup> "big data" is high-volume, high-velocity and high-variety information assets that demand costeffective, innovative forms of information processing for enhanced insight, and decision making.

The *volume* of data continues to increase astronomically. The *velocity* of data adds an interesting dimension. Data is coming to us at increasing speed, number and frequency of transactions. The velocity of big data needs to be understood, prioritized and synced up with our strategic needs both clinically and operationally. It is said that *variety* is the spice of life, and this is true for data too. We today have at our disposal an amazing array of data types, and across industries, amazing insights are being derived from text, geo-locations or log files of various sorts. Data is now a rich organizational asset – a healthcare entity's "natural resource" that is waiting to be tapped.

Data pundits are also focusing on additional V's in defining big data. One such is *verification*. The argument is that with big data comes such a diverse range of data that quality of data and well as security become big focal points for verification of the data. As we extract, transform and load (ETL) the big data into enterprise data warehouses, verification of quality and compliance becomes an important aspect of big data.

Value is also cited sometimes as an additional V. Deriving value out of insightful analytics that could impact care processes and outcomes is an important and often the most critical goal of efforts related to big data. The promise of the value, especially in healthcare, is one of a tremendous wave of innovation, progress, growth, and new care models – from the insights derived out of big data.

### Data rich, information poor

With the proliferation of information silos, the challenges associated with these silos of systems are many. In healthcare, clinical information systems are often siloed and often do not communicate or interact with each other. It seems that in today's environment, we are data rich yet information poor. As clinicians, we often find ourselves playing the role of a detective, navigating from one clinical system to another, piecing information together around our patients. The healthcare industry has made strides with data interoperability, but challenges still remain.

Whether in healthcare or across other industries too, data can be categorized as structured data (eg, laboratory data), unstructured data (eg, post-op notes and patient discharge summaries), imaging data (eg, computed tomography imaging studies), and streaming data (e.g. electroencephalography data). New technology has emerged that discovers, indexes, searches, and navigates diverse sources of data. We have massive amounts of databases and data repositories that feed into data warehouses. Add to this the massive amounts of data being generated from next generation genomic sequencers and a sea of electronic devices and equipment, we can very quickly see that the game of "connect the dots" becomes increasingly more complex.

#### **Data-driven insights**

The ability to run deep analytic queries on huge volumes of structured and unstructured data is a big data problem. It requires massive parallel processing data warehouses and purpose-built appliances for deep analytics, as well as capabilities around natural language processing (NLP) that are continuing to be perfected. Big data isn't just about data that is at rest - there is a significant amount of big data that is also in motion. Streaming data represents an entirely different big data problem – the ability to quickly analyze and act upon data while it's still moving. There has been much progress in this area, and the possibility of correlating data elements such as hours (or months) of live waveforms from the Intensive Care Unit (ICU) with other types of data across the healthcare enterprise is an exciting one.

There is a merging of traditional and big data approaches to handling these data elements. If the

33

# ENTERPRISE IMAGING

traditional approach was structured and repeatable analysis, and big data approach is one of iterative and exploratory analysis. Big data then delivers a fluid platform to enable creative discovery, and the user (eg, clinician, administrator or analyst) explores the facets and dimensions around the many ways intelligent insights could be asked or derived.

The opportunity at hand is to be able to scan across these massive stores of data, and connect them with other types of data that may be able to provide new insights and meaning. Correlating clinical data with cost, outcomes and performance data, and then tying these to evidence-based guidelines and clinical best practices could reveal entirely new insights and opportunities to continue to push the needle forward with newer care models.

The massive computational power required to crunch through big data is now becoming increasing accessible because of cloud technologies. As an example, IBM's Watson previously required racks of massive and expensive servers to crunch through truly big blobs of data. Today however, much progress is being made to make these services available through a cloud based deployment, with Watsonas-a-service<sup>5</sup> being a real offering. This is true for a variety of other big data platforms too such as Hadoop and Amazon Web Services (AWS) and this will continue to be a growing trend. What this does too is to bring along with it all of the benefits of cloud based deployments and access such as rapid elasticity, on-demand self-service and high performance with a lower initial capital investment.

In many ways, the essence of the conversation around data management has shifted with the availability of big data tools and capabilities. The debate today is less about whether we can afford to *store* information and more about whether we can actually afford to *throw it away*. The focus is moving from processing volumes of data that perhaps were not previously practical to store to dealing with massive amounts of data at a time, detecting insightful metrics, and responding quickly. The focus is also around data integration and data governance—managing data quality, security and information lifecycle management—all enabling much more meaning levels of actionable insights from the data at hand.

#### Big data and healthcare: Onwards and upwards

Big data enabled by cloud technologies could provide us new insights—clinically, operationally,

and in research—even as we focus on diagnostics across complex and challenging chronic illnesses and look at populations of patients in an increasingly dynamic and cost-conscious healthcare environment that is all about accountable care and value base care. Clinicians and researchers are dreaming up scenarios such as detecting patterns to a complex treatment regimen and contextualizing this to the specifics of the patient's clinical presentation and genomic as well as phenomic data, and then circling this back to metrics such as hospital readmissions, cost, and satisfaction.

As we expand our focus from disease to wellness, the opportunities to hone in on actionable insights leveraging analytics tools on big data platforms could shift the landscape from cost acceleration to cost deceleration. Consumer engagement could become vastly more meaningful with these deep analytic capabilities that could pull data across diverse data sources including from social media, wearable devices, and search trends, and allow for novel ways to incentivize healthy behavior and target high risk, high cost segments of the population.

McKinsey & Company predicts that if healthcare entities in the United States were to use big data creatively and effectively to drive efficiency and quality, the healthcare sector could create more than \$300 billion in value every year. As we get a better handle on the volume, velocity, and variety of data across healthcare, what we do with it to derive value then just becomes limited only by our imagination. Managing, analyzing, visualizing, and extracting useful information is becoming increasingly sophisticated yet doable. The road ahead is an exciting one, and we will only be limited in what we do with the data by the boundaries of our imagination.

#### REFERENCES

1. Shrestha RB. Accountable care and value-based images: Challenges and opportunities. Appl Radiol. 2013;4:19-22.

2. Big data: The next frontier for innovation, competition, and productivity. s.l. McKinsey Global Institute. June 2011.

 Hilbert M, Lopez P. The world's technological capacity to process information. University of Southern California. Accessed February 11, 2014. http://martinhilbert.net/WorldInfoCapacity.html.

4. Laney D. *3-D Data Management: Controlling Data Volume, Velocity and Variety.* Meta Group (Gartner).February 6, 2001.

5. Watson as a service: IBM preps AI in the cloud. *InfoWorld*. Updated November 15, 2013. Accessed February 12, 2014. http://www.infoworld.com/t/cloud-computing/watson-service-ibm-prepsai-in-the-cloud-230901.