A Look at Challenges and Opportunities of Big Data Analytics in Healthcare

Raghunath Nambiar
Cisco Systems, Inc.
San Jose, CA 95134, USA
rnambiar@cisco.com

Adhiraaj Sethi
Cisco Systems, Inc.
Herndon, VA 20171, USA
adysethi@cisco.com

Ruchie Bhardwaj
Cisco Systems, Inc./University of Southern California
San Jose, CA 95134, USA
rbhardwa@usc.edu

Rajesh Vargheese
Cisco Systems, Inc.
Austin, TX 78759, USA
rvarghee@cisco.com

***Abstract*— Big Data analytics can revolutionize the healthcare industry. It can improve operational efficiencies, help predict and plan responses to disease epidemics, improve the quality of monitoring of clinical trials, and optimize healthcare spending at all levels from patients to hospital systems to governments. This paper provides an overview of Big Data, applicability of it in healthcare, some of the work in progress and a future outlook on how Big Data analytics can improve overall quality in healthcare systems.**

Keywords- Big Data, Healthcare

I. INTRODUCTION TO BIGDATA

Big Data refers to large, complex datasets that are beyond the capabilities of traditional data management systems to store, manage, and process in a timely and economical manner. Often in petabytes, structured, semi-structured, and unstructured, Big Data creates challenges in data capture, transfer, encryption, storage, analysis, and visualization [1]. Big Data is a disruptive phenomenon, still in the early stages of adoption for many sectors, but it is very clear that harnessing its capabilities can provide compelling benefits.

Web 2.0 companies were most successful in monetizing Big Data. Many Big Data related innovations were also developed by them, contributing to the industry a growing collection of open source technologies that have dramatically changed the culture of collaborative software development and the scale and economics of hardware infrastructure [2].

Industry analyst reports show that Big Data and analytics can generate significant financial values in many vertical markets, including healthcare, finance, retail, environmental research, genomics, and biological and life science research. Big Data market is estimated to grow 45% annually and expected to be over $25 billion by 2014 [3][4]. See Figure 1.

Considering the fact that healthcare spending is one of the biggest financial challenges of the U.S., optimizing it while improving the quality of care is extremely important for patients to health care providers to government agencies. Long term fiscal challenges for federal and state governments include the growing costs of Medicare, Medicaid, and other public health programs, which are often hampered by lack of access and timeliness of the data that reside across various computer systems silos [5]. According the Harvard School of Public Health publication entitled *The Promise of Big Data*, petabytes of raw information could provide clues for everything from preventing tuberculosis to shrinking health care costs—if we can figure out how to apply this data [6]. Improving the care of chronic diseases, uncovering the clinical effectiveness of treatments, and reducing readmissions are expected to be top priority use cases for Big Data in healthcare [7].

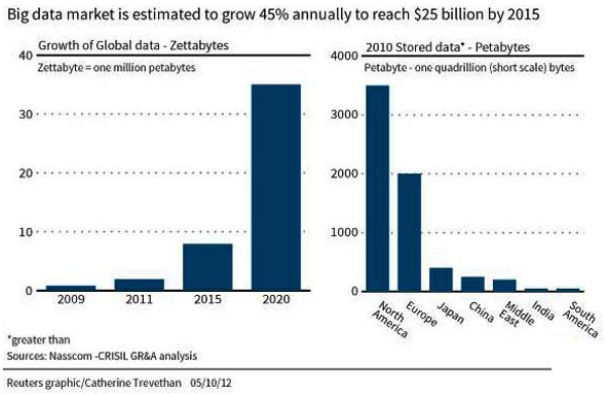

Figure 1. Big Data Growth

.

II. HEALTHCARE INDUSTRY OVERVIEW

The healthcare industry is one of the most critical industries. It is also one of the world's largest and fastest-growing industries [7][8].

The U.S. healthcare industry is evolving through significant challenges in recent times. Some of the key challenges include, but are not limited to: escalating costs of care delivery, raising patient volumes with the passage of the Affordable Care Act, an aging population with a high incidence of chronic disease, substantial shortage of medical professionals, and penalties for readmissions and significant reductions in reimbursement.

In most countries, healthcare costs account for a good percentage of its economy. For example, the cost of healthcare spending in 2012 in the U.S. was estimated at $2.6 trillion, approximately 17.6% of gross domestic product (GDP) [9]. See Figure 2. In the future, the cost of healthcare is projected to continuously increase. For example, estimates

978-1-4799-1293-3/13/$31.00 ©2013 IEEE

put U.S. healthcare costs at $4.1 trillion in 2016. Individual segments within healthcare industry also follow the same trend. A report by IMS Research estimates that spending for pharmacy drugs will increase to $1.2 trillion by 2016 [9].

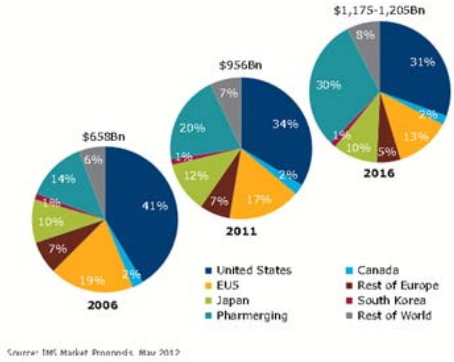

Figure 2. Healthcare spending by geography

While the healthcare industry is very critical and vast, it is unfortunately highly inefficient. According to a report from the Institute of Medicine, one third of the spending in healthcare (approximately $750 billion) today is wasted and not contributing to improving healthcare outcomes. A further breakdown analysis of the waste in healthcare in the same report from the Institute of Medicine indicates many areas such unnecessary services, administrative waste including unproductive and duplicate documentation, inefficiently delivered services, high prices, fraud and missed prevention opportunities [10].

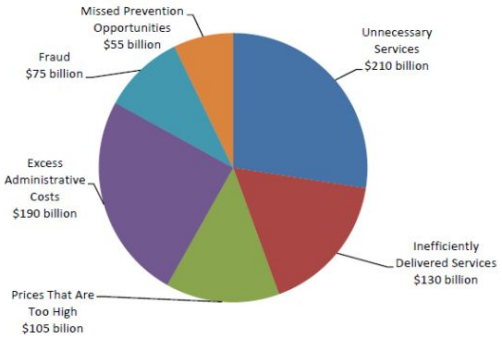

Figure 3. Waste in Healthcare spent

To understand the sources of inefficiencies in healthcare, it is extremely important to understand how data, processes, people, and devices enable care coordination. Lack of coordination and standards result in higher cost and poor outcome. It is estimated that in one out of five elderly patients are readmitted within 30 days of discharge for no known reason. See Figure 4.

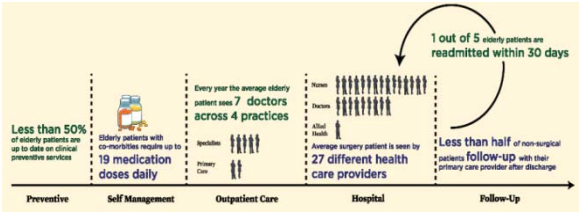

Figure 4. Example of inefficiencies in healthcare

Insurance companies and governments reimburse for procedures performed rather than outcomes achieved. Patients bear little responsibility for the cost of the health care services they demand

Reforming healthcare to improve the quality and care is essential to address efficiencies in the system, increasing costs, poor quality, and increasing numbers of people without health insurance coverage. Major healthcare reforms are being debated in the U.S. and around the world. The focuses of healthcare reforms include:

- Improving quality, safety, efficiency, and reducing health disparities
- Engaging patients and family
- Improving care coordination and public health
- Maintaining the privacy and security of patient health information
- Enhancing clinical outcomes
- Improving population health outcomes
- Increasing transparency and efficiency
- Empowering individuals
- Generating more robust research data on health systems

As the healthcare industry focuses on improvements, Big Data Analytics can play a significant role in addressing the challenges of mining the relevant information at the point of care to enable safer, faster, more cost effective, and more efficient medical practices.

III. BIGDATA IN HEALTHCARE

The Big Data revolution has begun for many industries. The healthcare industry has been playing catch up and has finally reached a consensus on the value of Big Data as a transformative tool. According to McKinsey & Company study entitled *The Big Data Revolution in Healthcare*, big-data revolution is under way in health care enable by increased availability of information. Major data sources include years of research and development data form pharmaceutical companies, Electronic Medical Records (EMRs) and Electronic Health Records (EHRs) from healthcare providers, and clinical trials data covered under public insurance programs from US government agencies. Recent technology advancements in hardware and software are making it easier to collect, transfer, storage, aggregate, and analysis data from multiple sources [24]. Yet there are privacy and standardization challenges that need to be addressed before we full adaption.

Overall, McKinsey & Company estimates that $300 billion to $450 billion can be saved in the healthcare industry from Big Data Analytics [12].

While healthcare industry is slowly realizing the value from Big Data, Kaiser Permanente, Blue Shield of California and AstraZeneca are few exceptions who have already implemented integrated analytics systems to improve efficiency [24].

Kaiser Permanente's HealthConnect system connecting all their facilities have improved outcomes in cardiovascular disease and achieved an estimated $1 billion in savings from reduced office visits and lab tests. Blue Shield of California, in partnership with NantHealth, has developed an integrated technology system to improve health-care delivery. AstraZeneca in partnership with HealthCore, a subsidiary of WellPoint, is conducting real-world studies to determine the most effective and economical treatments for some chronic illnesses and common diseases [24].

Some argue that the key to making Big Data a greater reality in healthcare is by integrating healthcare provider systems with insurance agencies and government agencies. If consensus cab be reached on privacy, legal, security and compliance issues the combination of the three agencies can optimize resources and analyze their shared information to spot trends, speed up recoveries, and perform both evidence based medicine and preventive medicine [12].

As new care models such as Accountable Care Organizations (ACO) emerge, the focus is on driving the wellness of and identifying the at risk population well in advance. Offering appropriate programs such as patient engagement and adherence is critical to the success of ACOs. Big Data patterns, trends, demographics, and claims can be used to predict future patterns. In a recent Healthcare Business News article entitled *Why one Medicare Pioneer ACO Succeeded in Saving Money*, the use of analytics powered by Big Data was clearly called out as a key initiative contributing to success. They are able to identify the risk for hospitalization using sophisticated algorithm [13].

Big Data analytics enables the capture of insights from data gathered from research, clinical care settings, and operational settings to build evidence for improved care delivery. Today, as indicated in the Institute of Medicine (IOM) report, these are critical gaps. There is a significant opportunity to improve the efficiencies in the healthcare industry by using an evidence-based learning model, which can in turn be powered by Big Data analytics [9]. As the focus shifts from cure to preventive health and as new technologies such as wearable sensors evolve as part of the Internet of Things (IoT), the volume of data in healthcare is expected to grow significantly and can provide a wealth of actionable information. The combined power of information from real-time devices, people, clinical systems, and historical population data makes Big Data a very helpful tool in improving the healthcare system.

IV. EXAMPLES: BIG DATA INITIATIVES IN HEALTHCARE

There are several initiatives utilizing the potential of Big Data in healthcare. Some of the examples are listed below:

Asthmapolis: The company has created a global positioning system (GPS) enabled tracker that monitors inhaler usage by patients, eventually leading to more effective treatment of asthma. By placing a small cap like device containing a GPS sensor on top of an inhaler, the company is able to gather usage information. When a patient is suffering from an asthma attack and is required to use his or her inhaler, the little device records the time and place that the inhaler was used and transmits this information to a web site. This data is then combined with information available through the Center for Disease Control (CDC). The CDC'S information includes facts such as what are the most prominent asthma catalysts present in the air for that location and where the highest pollen count is located. The end product is the ability to show asthmatics how other people with similar reactions to specific allergens reacted and if taking an inhaler was necessary. A patient can become more aware when making decisions to travel and be more aware if they should prepare for a potential asthma attack. The combination of this large amount of data can also allow a physician to create a personalized plan for the patient, targeting specific times of the day when the risk of an asthma attack is higher and responding by either increasing or decreasing the dosage [14].

Battling the Flu: Big Data analytics has become a weapon for CDC to combat influenza, which claims millions of lives a year. Every week, the CDC receives over 700,000 flu reports. These reports include the details on the sickness, what treatment was given, and whether not the treatment was successful. The CDC has made this information available to the general public called *FluView*, an application that organizes and sifts through this extremely large amount of data to create a clearer picture for doctors of how the disease is spreading across the nation in near real-time. Along with providing the precise locations of patients combating the flu, the application provides caregivers with information such as what vaccines combat which strain of the flu and which antiviral medications can aid the patient in their recovery [25].

The applications of Big Data for combating the flu don't stop at what a doctor can do for a patient; there is increasing development of applications that are meant to assist patients in their own recovery from the flu. An application called *FluNearYou* created by the Skoll Global Threats Fund and the American Public Health Association, asks users to input their symptoms in the application before they actually fall ill with the sickness. Thus, by gathering data when people start falling sick, a map of flu activity is generated and could allow users to prepare themselves and prevent exposure to the bug. Another tool, *Germ Tracker*, is similar to *FluNearYou* but instead pulls data exclusively from social media. The benefit of this is that analysts have access to the data from thousands of cases that have gone unreported to doctors. A disadvantage is that in some cases a person posting on social media may be incorrect about their self-diagnosis. Another application is *Help, I Have the Flu*, designed by the pharmaceutical company Help Remedies. This application also takes advantage of social media and searches for the person responsible for passing the flu to

individuals by scanning through updates for key words such as “flu” and “cough” [15].

GNS Healthcare and Aetna: GNS Healthcare, a Big Data analytics company, has come together with the health insurance company Aetna to help combat people at risk or already with metabolic syndromes. GNS has developed a technology known as Reverse Engineering and Forward Simulation (REFS) that will be put to work on the data of Aetna insurance subscribers. Essentially, the technology will search for the presence of five warning signs: large waist size, high blood pressure, high triglycerides, low High-density Lipoprotein (HDL), and high blood sugar. A combination of any three of these lead to the conclusion that the patient is suffering from the condition. Two or less suggest that the patient may be at high risk of suffering from the condition. Different combinations of these warning signs lead to different conclusions. For example, low HDL and high triglycerides is a sign of high-risk hypertension within the next 12 months while high blood sugar and high triglycerides could lead to a completely different conclusion [16].

Diabetes and Big Data: Diabetes patients can also benefit from the Big Data revolution in health care. A company named Common Sensing has produced GoCap, a cap for prefilled insulin pens that can not only record the amount of insulin administered daily but also the specific times the dosages were administered. This information is then transmitted to either a mobile phone where an application records this data or to a connected glucometer. This data is then easily available to healthcare professionals and allows them to identify problems before they become severe and to tweak dosages if required [17].

Another technology available at the intersection of Big Data and diabetes is a service offered by Allazo Health. Their system, *AllazoEngin,* is designed to improve on medication adherence programs by using predictive analytics. For example, predict what interventions are mostly likely to work for that patient based on what interventions already worked for other patients with similar demographics, behavioral profiles, and medical history [10].

*23andMe:*A privately held personal genomics and biotechnology company, provides rapid genetic testing. The company is named for the 23 pairs of chromosomes in a normal human cell. Their personal genome test kit was named "Invention of the Year" by *Time* magazine in 2008 [16]. Mining DNA data from thousands of others and comparing it with their own genomic data will help understand genetic health risks and inherited conditions.

USC Medical Monitor: Computer scientists at the University of Southern California (USC) are teaming up with neurologists, kinesiologists and public health experts to fight against Parkinson’s disease. The team uses various devices to track the movement of the patient and gather large amounts of data including data from 3D sensors of Microsoft Kinect, patient’s smartphone, and additional body sensors. Then the data is fed into an algorithm that analyzes the data to identify any significant changes in movement. Cesar Blanco, the head of the project, stated, “Our system will allow patients and their caregivers to monitor disease progression and the effectiveness of treatments in real-time.” For example, if the system identifies a decreased range of motion or a decreased flexibility, an alert is sent to both patient and caregiver. From there, the application will advise the patient what exercises to do that will strengthen their body and the caregiver can prescribe different medications or change the dosage of a medication. The team hopes to extend this technology beyond Parkinson’s Disease in the future [19].

MD Anderson’s Moon Shots Program to Fight Cancer: Just as President John F. Kennedy in 1962 presented the challenge of putting a man on the moon before the end of 1960s, MD Anderson president Ronald DePinho, M.D., has challenged scientists and clinicians to rapidly and significantly reduce mortality in several major cancers [20].

The Moon Shots Program aims to improve survival for several types of cancers. It is initially targeting eight cancers, selected based on rigorous criteria: the current state of scientific knowledge across the continuum from cancer prevention to survivorship; the strength and breadth of the assembled teams; and the potential for measurable success in reducing deaths. The selected cancers are: Triple-negative breast and ovarian, (which are linked at the molecular level, Leukemia (AML/MDS), Leukemia (CLL), lung, melanoma, and prostate.

The Moon Shots Program is enabled by several cross platform collaboration efforts between various industries and cancer researchers. Big Data and Massive Data Analytics are a major part of this program. A Big Data platform enables centralization, integration, and secure access of patient and research data and analytical results. A Massive Data Analytics infrastructure enables complex analytics and clinical decision support systems using integrated patient information, including clinical and research data. In addition, an Adaptive Learning in Genomic Medicine framework for bringing clinical medicine and genomic research together enables rapid learning to improve patient management using Clinical Genomics, Omics-Bioinformatics, and Massive Data Analytics platforms within the Big Data environment [20].

White House Big Data Research and Development Initiative: Announced by the Obama Administration in 2012, this initiative aims to make the most of the fast-growing volume of digital data by investing in advanced technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data and harness these technologies to accelerate the pace of discovery in science and engineering, national security, and to transform teaching and learning. This initiative include a joint solicitation supported by the National Science Foundation (NSF) and the National Institutes of Health (NIH) to advance the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large and diverse data sets particularly in imaging, molecular, cellular, electrophysiological, chemical, behavioral, epidemiological, clinical, and other data related to health and disease.

In addition, the NIH Big Data to Knowledge (BD2K) Initiative is designed to help biomedical scientists fully utilize Big Data being generated by research communities. BD2K Centers of Excellence will address these issues by

developing new approaches, methods, software, tools, and related resources and by providing training to advance Big Data science in the context of biomedical research. BD2K Centers are expected to be large-scale efforts that take on challenges that are not feasible to address with the standard NIH Research Project Grant (R01) program. The Centers will form a BD2K Center Consortium in which each is required to interact and collaborate with other centers in the consortium. They are also encouraged to interact and collaborate with other domestic or international efforts in Big Data science. The BD2K Centers will work in areas spanning Big Data science, producing tools and resources from early-stage to mature development for the biomedical research community. The NIH would like the overall BD2K Center program to covers four areas: Collaborative environments and technologies, data integration, analysis and modeling methodologies and computer science and statistical approaches [21].

V. INFRASTRUCTURE FOR BIG DATA IN HEALTHCARE: CISCO SOLUTION

As a worldwide leader in networking, Cisco is well positioned to help improve the future of healthcare through networked technologies that transform how people connect, access and share information, and collaborate. Cisco technologies can benefit all stakeholders, from patients to medical providers, payers, and life sciences organizations. Cisco Connected Health solutions enable care team collaboration and business efficiency while promoting a higher quality of care.

Cisco helps simplify healthcare communications using a network of interoperable technologies that better connect patients with medical providers, payers, and life science organizations. Cisco solutions link critical information, people, and knowledge to help improve the healthcare experience. Our vision is a Connected Health ecosystem that helps to shape a world of health without boundaries.

The Cisco Architecture for Connected Health was developed specifically for healthcare providers to help them build a robust, highly secure, and scalable infrastructure for the delivery of healthcare service by utilizing the guidelines and best practices from Cisco's health care experience which is focused on the hospital data center, unified workspace, and interoperability for healthcare. See Figure 5.

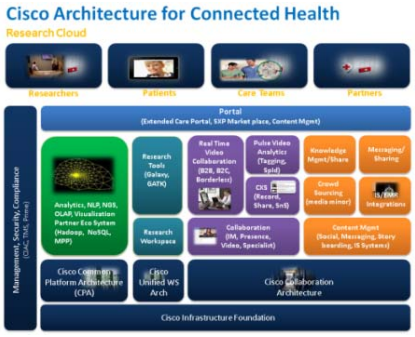

Figure 6. Cisco's Architecture for Connected Health

Cisco also builds an innovative family of unified computing products. This enables the company to provide a complete infrastructure solution including compute, storage, connectivity and unified management for Big Data applications that reduce complexity, improves agility, and radically improves cost of ownership. Cisco's Big Data solution can meet standards like the Payment Card Industry (PCI), Data Security Standard (DSS), and Health Insurance Portability and Accountability Act (HIPAA) important for the healthcare industry.

Core to Cisco's Big Data architecture is the Cisco Unified Computing System™ (Cisco UCS®) Common Platform Architecture (CPA) for Big Data [22], a complete infrastructure solution including compute, storage, networking, and unified management designed to meet a broad range of applications demands.

With recently announced support for Intel Distribution for Apache Hadoop [23] the architecture enables advanced access control, and encryption of data-in-transit and data-at-rest, critical for Big Data deployments in healthcare.

VI. FUTURE OUTLOOK

Today the healthcare industry is just beginning to understand all the innovative things that can be done with Big Data. The intersection of data from multiple sources, tools, and technologies will promote informative extrapolations of Big Data, allowing the information to generate new and innovative solutions to healthcare.

The healthcare system today is on a trajectory that is unsustainable. The bulk of costs in the current system are due to patients having chronic care diseases. Thus a focus in the future should be on preventive care as well as population health management and overall wellness. With Big Data, health management of a population can be understood better.

Personalized medicine is a being promoted as the future of the healthcare industry. Today, medicines are made for the masses not for the individual. Going forward, with the help of Big Data, more personalized medicines that uses patient-specific data such as genomics and proteomics can be created based on the profiling of similar patients and their responses to such approaches.

With the increased adoption of social media and mobility patients today are becoming more and more knowledgeable about the options available to them. This is made possible by tools and technologies that allow them to access the data and reports but in the backend powered by Big Data platforms that intern do massive searches, aggregations, and pattern recognitions. New data sources and analytics technologies are expected to emerge in the near future that will change the way we practice medicine.

Greater collaboration across industries and between the public and private sectors can improve efficiencies in healthcare system. By integrating data from various sources can build predictive models that can lower overall cost and improve quality of care significantly, and accelerate the shift to evidence-based and preventive-based medicine. With

better analytics, the risks will at least be known and the decisions can more informed and quicker to make.

VII. ACKNOWLEDGE

The authors thank Dr. Anakha Nambiar for her comments and feedback.

REFERENCES

[1] Big Data, Wikipedia: www.wikipedia.org

[2] Raghunath Nambiar: Data Management – A Look Back and a Look Ahead, WBDB 2012

[3] IDC, Digital Universe in 2020

[4] Nasscom-CRISIL GR&A Analysis, Reuters

[5] Eric Dishman, How Big Data can revolutionize health care:, Politico

[6] The Promise of Big Data, Harward School of Public Health, 2012

[7] Allen Bernard: Healthcare Industry Sees Big Data As More Than a Bandage

[8] Industry Classification Benchmark: https://en.wikipedia.org/wiki/Industry_Classification_Benchmark, 2013

[9] The Global Use of Medicines: Outlook Through 2016,IMS Institute, 2012

[10] Peter Groves, Basel Kayyali, David Knott, Steve Van Kuiken: The 'Big Data' Revolution in Healthcare: Center of US Health System Reform Business Technology Office, 2013

[11] Devon M Herrick, Linda Gorman, John C. Goodman: Health Information Technology: Benefits and Problems: National Center for Policy Analysis, Policy Report No. 327: 2010

[12] Chris Poelker: Centralizing healthcare big data in the cloud: http://blogs.computerworld.com/cloud-computing/20488/centralizing-healthcare-big-data-cloud: 2012

[13] Jessica Zigmond: Why one Medicare Pioneer ACO succeeded in Saving Money, Healthcare Business News

[14] Shel Israel: Contextual Health vs The Elephant in the Hospital, 2013

[15] Mike Wheatley: Healthy Big Data: Stopping the Spread of Flu, 2013

[16] Arundhati Parmar: Want to know if you will develop high blood pressure next year? With big data magic you can, 2012

[17] Stephanie Baum: A remote monitor embedded in insulin pen caps could help personalize diabetes treatment, 2013

[18] Cliffard Jones: 2013 Data Design Diabetes Innovation Challenge, 2013

[19] Marc Ballon: Medical Monitor: Los Angeles: USC Trojan Family, 2013

[20] MD Anderson's Moon Shot to Fight Cancer. http://www.cancermoonshots.org/about

[21] White House Press Release. Obama Administration Unveils "Big Data" Initiative, 2012

[22] Raghunath Nambiar: Cisco UCS Common Platform Architecture (CPA) for Big Data : http://blogs.cisco.com/datacenter/cpa/

[23] Cisco UCS Common Platform Architecture (CPA) for Big Data, enhanced with encryption at rest: http://www.cisco.com/en/US/docs/unified_computing/ucs/UCS_CVDs/Cisco_UCS_CPA_for_Big_Data_with_Intel.html

[24] Basel Kayyali, David Knott, Steve Van Kuiken: The big-data revolution in US health care: Accelerating value and innovation, McKinsey & Company

[25] Julie Bort: How The CDC Is Using Big Data To Save You From The Flu, Business Insider