insideBIGDATA

InsideBIGDATA Guide to

Healthcare & Life Sciences

by Daniel Gutierrez

BROUGHT TO YOU BY DELL EMC | intel

Big Data and Analytics for Healthcare and Life Sciences – An Overview

The healthcare and life sciences industries historically have generated vast amounts of data. These large volumes of data hold the promise of supporting a wide range of medical and healthcare tasks, including clinical analytics and decision support, patient profiling, disease surveillance, regulatory and compliance requirements, scientific research, and many others. Data in healthcare and life sciences is expected to grow exponentially in the coming years and will be beyond the capability of the traditional methods of data management and data analytics.

It is vitally important for organizations in these industries to acquire the available infrastructures, methodologies, and tools to leverage this vast amount of data effectively to ensure the highest possible standard of patient care, as well as risk significant revenue and potential profits. This technology guide provides an overview of the utilization of big data technologies as an emerging discipline in healthcare and life sciences. It explores the characteristics of this business strategy and the benefits of leveraging big data technologies within these sectors. It also touches on the challenges and future directions of big data and analytics in the healthcare and life sciences industries.

Despite the fact that some data in the healthcare sector is still stored in hardcopy form, most is in electronic form. One issue, however, is that this data is now stored in electronic silos with more and more data produced every day from new devices. Big data in the healthcare industry promises to support a diverse range of healthcare data management functions, however the industry is still in the early stages of getting its feet wet in the large scale integration and analysis of big data.

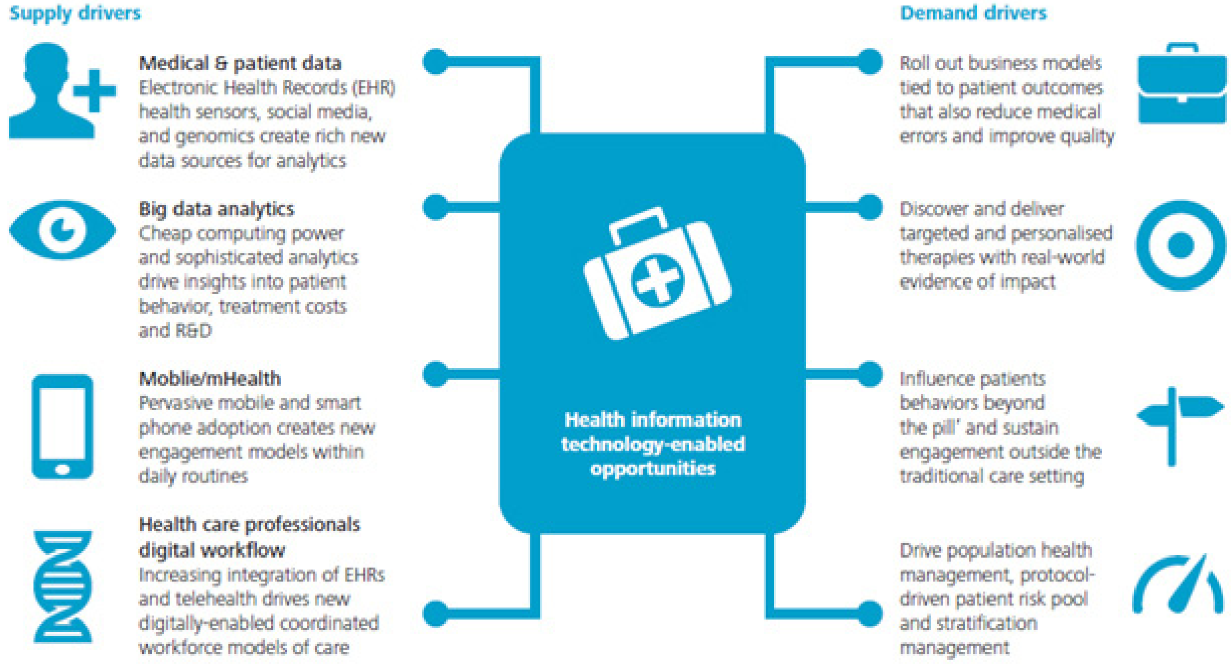
Life sciences research continues to evolve rapidly in conjunction with an increasing focus on analytics and the more effective use of data. The race to understand patients and diseases at the molecular level to achieve precision medicine is fueling this shift. The figure on page three encapsulates many of the demand drivers coupled with new business models where outcomes and real-world data are providing health data and transforming what is possible. It's clear that the race to find the cause and subsequent treatments and cures is paramount.

Contents

Big Data and Analytics for Healthcare and Life Sciences – An Overview 2
Use Case Examples Abound 4
Challenges for Adopting Big Data and Analytics in Healthcare and Life Sciences 6
The Rise of Deep Learning 8
Distributed Systems – the Key to Success 9
- Hadoop Use Cases 9
- Spark Use Cases 10

The Impact of IoT on Healthcare and Life Sciences 11
The Convergence of Big Data and HPC 12
Case Studies: Dell EMC Focused Customer Use Cases 13
- Healthcare Provider 13
- TGen 13
- Arizona State University 14
- Transforming Patient Care at the University of Iowa Hospital and Clinics 14
- Centegra Health System 14
- Large Healthcare Organization 14

Summary 15



Source: *Healthcare and Life Sciences Predictions 2020: A bold future?*, U.K. Centre for Health Solutions, 2014

The New York based research and consulting firm, *Institute for Health Technology Transformation* estimated that in 2011, the US Healthcare industry generated 150 exabytes of data — enough to copy all of the printed materials in the Library of Congress 15 MILLION times over. This data was mostly generated by patient care, record keeping, and various regulatory requirements. Since then, there has been an exponential increase in data which has led to an expenditure of $1.2 trillion towards data solutions in the healthcare industry. Healthcare expenses in the U. S. now represent 17.6 percent of GDP — nearly $600 billion more than the expected benchmark for a nation of its size and wealth. McKinsey & Company projects that the use of big data solutions in healthcare can reduce the healthcare data management expenses by $300 billion – $500 billion.

Big data in healthcare originates from large electronic health data sets — these data sets are very difficult to manage with the conventional hardware and software. The use of legacy data management methods and tools also makes it difficult to usefully leverage all this data. Big data in healthcare is an overpowering concept not just because of the volume of data but also due to the variety of data types and the velocity at which healthcare data needs to be managed. Furthermore, the sum total of data related to the patient and their well-being constitutes a rising problem in the healthcare industry.

Big data technologies allow leading healthcare and life sciences organizations to address and overcome a wide variety of business and clinical challenges such as improving patient safety, reducing 30-day hospital readmissions, and enhancing drug discovery by harnessing the power of their data with big data analytics solutions. These technologies enable organizations to import, unify and analyze information from traditionally isolated data silos such as patient electronic health records (EHR), claims information from payer organizations, even data outside the traditional healthcare context such as socio-economic and patient generated data in a scalable, cost-effective manner. Importantly, these solutions comprehend both structured and unstructured information and help organizations move from historical reporting to real-time predictive analytics.

One class of big data technology that is particularly useful for healthcare and life sciences is the data lake. Dell EMC has specific experience in bringing data lake technology to the healthcare and life sciences community. The instruments used to gather data is going through its own "Moore's Law" of getting cheaper and faster. A data lake is all about volume of data arriving from new medical and research devices, but it's also about variety — not just patient and patient genetic data, but also patient molecular data, functional data, image data, etc. The data lake is also about a security layer on top. A data lake accelerates discovery by placing all this data in one place, along with data not even conceived of before. New tools and algorithms will need to be discovered to fully take advantage of a data lake.

> With the power of big data and data science, we are one step closer to a world where genetic diseases are more effectively managed and more frequently cured, changing patient lives forever.

Big data technologies enable researchers to perform analyses and make informative, actionable decisions that are driving real change in the treatment of rare genetic diseases — making it easy for biologists to identify genetic disease markers and assess drug efficacy when visualizing cell data. This change also allows for faster time-to-value for pharmaceutical companies as well as a shorter path to patient benefits, e.g. identification, diagnosis and predictive analytics in action. With the power of big data and data science, we are one step closer to a world where genetic diseases are more effectively managed and more frequently cured, changing patient lives forever.

Use Case Examples Abound

There are an increasing a number of important use case examples where combining big data technology with healthcare and life sciences has become meaningfully beneficial:

- **Genome processing and DNA sequencing –** there is tremendous growth occurring in the genomics sequencing market as evidenced by data volume increases produced by DNA sequencers and in the number of individuals being sequenced (even though much of the data coming out of a sequencer is not actionable and not usable in the EHR). Additionally, "medicine" starts after the VCF (variant call format) file is annotated and is part of the Interpretation phase, which could happen in part in Hadoop, although there are other options like SAP HANA and Microsoft Analytics Platform System (APS). Using SAP® Foundation for Health™, built on SAP HANA, helps turn big data into smart data, adding value for healthcare organizations, and life sciences companies.

> Researchers plan to build instruments that will monitor the activity of hundreds of thousands and perhaps 1 million neurons, taking 1,000 or more measurements each second.

- **Neuroscience –** the U.S. based BRAIN Initiative uses big data technologies to map the human brain. By mapping the activity of neurons in the brain, researchers hope to discover fundamental insights into how the mind develops and functions, as well as new ways to address brain trauma and diseases. Researchers plan to build instruments that will monitor the activity of hundreds of thousands and perhaps 1 million neurons, taking 1,000 or more measurements each second. This goal will unleash a torrent of data. A brain observatory that monitors 1 million neurons 1,000 times per second would generate 1 gigabyte of data every second, 4 terabytes each hour, and 100 terabytes per day. Even after compressing the data by a factor of 10, a single advanced brain laboratory would produce 3 petabytes of data annually.

- **Personalized treatment planning –** a way to customize treatment for a patient to continuously monitor the effects of medication. The dose can be adapted or the medication changed based on how the medication is working for that particular individual. This analysis can be applied at the individual level and is tailored to each patient's specific needs. But personalized medicine goes far beyond monitoring the effects of medication. Precision medicine is defined as an emerging approach for disease management and prevention that takes in to account individual variability of genes, environment and lifestyle for each person. Precision medicine looks to move the needle from reactive medicine to proactive medicine. It's a move away from drugs and treatments for the broad population to a more precise, individualized treatment plan. A recent study from the University of California, San Diego, found that patients who are treated utilizing a more precise plan or more personalized plan saw improved wellness periods of nearly 30% versus the non-precision medicine group.

Assisted diagnosis is accomplished using expert systems that contain detailed knowledge of conditions, symptoms, medications and side effects. Bringing together individual data sets into big data algorithms provides more accurate insights.

- **Assisted diagnosis –** being able to access a broad combination of knowledge across multiple data sources aids in the accuracy of diagnosing patient conditions. Assisted diagnosis is accomplished using expert systems that contain detailed knowledge of conditions, symptoms, medications and side effects. Bringing together individual data sets into big data algorithms provides more accurate insights.
- **Using predictive analytics** to help accurately predict discharge dates, help identify patients at high risk for readmission, and help surgical teams keep patients safe by reducing surgical site infections by 58 percent while decreasing the cost of care (as reported by the University of Iowa).
- **Using machine learning tools** to circumvent diagnosis codes in EHRs that are fraught with accuracy problems, to automate the detection of both false positives and missing codes in patient charts. Use text data from millions of doctors' notes to train a machine learning classifier to pick out heart failure patients based on everything in their charts, not just their diagnosis codes.
- **Creating tools that integrate genetic data and accelerate its use** at point of service. Integration, however, is not an easy task since genetic data must follow the patient over their lifetime throughout their episodes of care.
- **Monitoring patient vital signs –** healthcare facilities are looking to provide more proactive care to their patients by constantly monitoring patient vital signs. The data from these various monitors can be used in real time and send alerts to nurses or care providers so they know instantly about changes in a patient's condition.
- **Evidence-based medicine –** involves making use of all clinical data available and factoring that into clinical and advanced analytics. The outcomes of this application of big data include improved ability to detect and diagnose diseases in their early stages, assigning more effective therapies based on a patient's genetic makeup, and adjusting drug doses to minimize side effects and improve effectiveness.
- **Using bioanalytics** platforms designed to improve the productivity of biologists.
- **Optimizing the EHR** through consolidation to reduce costs and increase efficiency.
- **Gaining insight,** preventing inefficiency, and adapting workflows for better healthcare.
- **Using sensing devices** for data collection as a revolutionary step forward in Parkinson's research.

New health plan solutions provide advanced business intelligence, analytics and strategic information management systems that can help derive meaningful insights from data to attract customers, as well as manage costs and risks.

- **Using collaborative analytics** via the cloud to personalize treatment plans.
- **Health plans** can move beyond traditional descriptive analytics and unwieldy data warehouse strategies. New solutions provide advanced business intelligence, analytics and strategic information management systems that can help derive meaningful insights from data to attract customers, as well as manage costs and risks.
- **Creating fraud detection solutions –** healthcare organizations need to be able to detect fraud based on analysis of anomalies in patient records, billing data or procedural benchmark data.
- **Imaging analytics –** opens up new diagnostic landscapes for interpreting x-rays, CAT scans, and MRIs which has largely remained under the responsibility of skilled clinicians who specialize in catching abnormalities and reporting on findings. In contrast, as computing power increases and analytics algorithms start to become intelligent enough to analyze patterns in digital images, these test results are taking on a whole new meaning for the diagnostic process.

Challenges for Adopting Big Data and Analytics in Healthcare and Life Sciences

The effective use of data analytics in healthcare has been hailed as a solution for saving time and dollars and improving patient outcomes for a healthcare organization. However, many health systems have a hard time capturing and using data from patients that can make a real impact on patient outcomes. Part of the issue lies in EHR data, which can provide an incomplete picture of patient behavior. EHR is the systematized collection of patient electronically-stored health information in a digital format. EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users. When big data can be used holistically to revamp healthcare processes including care coordination, patient relationships and financial services, the results can save organizations time and money.

According to research by healthcare technology provider Evariant, Inc., the effective use of big data technologies in healthcare could save the industry $300 million a year. Analysis of real-time data saved one hospital $850,000 in overtime costs alone, using more intelligent discharge planning, disease management, quality assurance, and performance reporting. However, it has been difficult for healthcare providers to integrate EHR data with clinical transcripts and other notations that would add context to patient care. According to a 2015 eHealth Initiative survey, only 17% of providers have been able to couple population health analytics with EHR data.

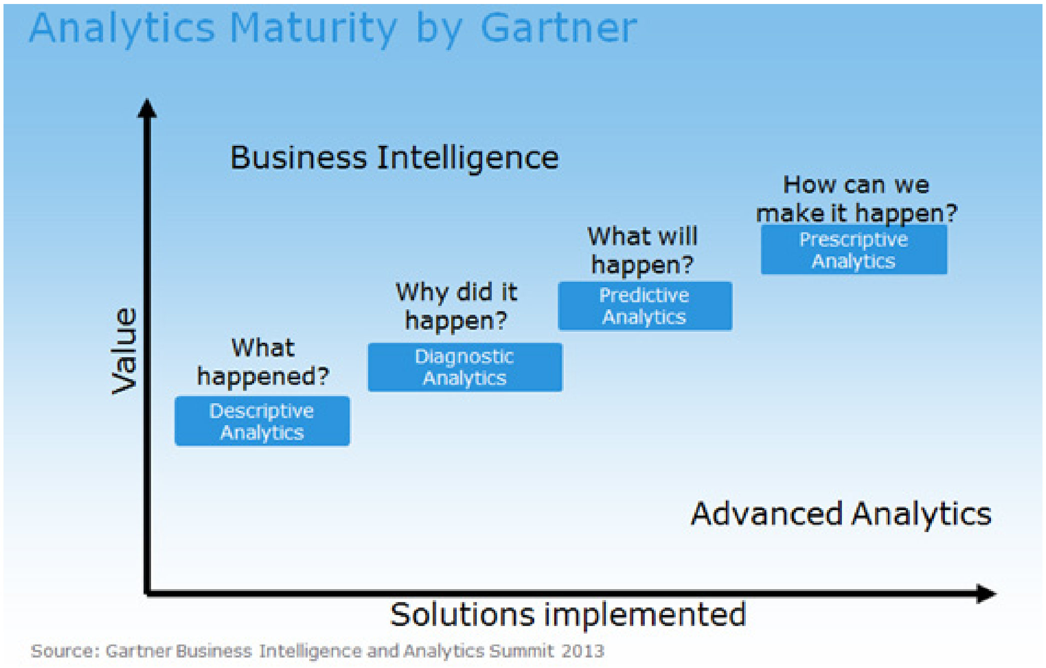
As life sciences organizations face growing challenges, being effective with data becomes essential for sustained success now and into the future. Those who understand how to manage both the internal and external data relevant to their products, markets and customers will create the opportunity for competitive advantage based on improved insight. If life sciences organizations are able to apply their acumen with big data and analytics to drive decisions and engage in smart collaboration, they will find order and opportunity where others see chaos.

With 80% of the healthcare data being unstructured and growing exponentially, it is a challenge for the healthcare industry to make sense of all this data and leverage it effectively for treatment courses, clinical operations, and medical research. It is extremely important for the big data healthcare companies to make use of the best-in-class technology that can leverage big data in healthcare effectively. Getting access to and using this unstructured data — such as output from medical devices, doctor's notes, lab results, imaging reports, medical correspondence, clinical data, and financial data — is an invaluable resource for improving patient care and increasing efficiency.

In order to address these needs, Dell EMC collaborated with Partners HealthCare to develop *The Partners Data Lake,* an agile data and analytics platform. The data lake will be used to inform medical science to help make research more efficient, improving the diagnosis and treatment of disease and ultimately making health care more cost effective. Leveraging the advanced medical diagnostics capabilities throughout the Partners HealthCare system, large amounts of data generated through separate research activities will now be brought together within the data lake to enable the medical community to explore and develop new insights into human disease to improve diagnostics, treatment and the lives of patients. The advent of personalized medicine and advancements in technology have changed the medical field. Treatments are accelerated thanks to personalized medicine and the advancing rate of medical research. At the same time, medical research has shifted towards a collaborative approach as technology advancements eliminate physical boundaries

In the next sections, we'll examine two areas of technology that have become quite prevalent for the healthcare and life sciences industries — deep learning and distributed computing architectures (e.g. Apache Hadoop, Spark), as well as other technologies like SAP HANA solutions for aggregating data types into a format that can easily be queried.

The overarching goal is to be able to engage the analytics maturity progression as depicted in the figure below. There is a tremendous opportunity to use predictive analytics to automate processes in healthcare. In the provider case, you can consult thousands of physicians and hundreds of thousands of cases, process real-time data, and learn the repeated patterns that allow you to make predictions about best courses of actions, best treatments, the probability and risk involved in sepsis, and more. We're looking at a revolution in the way healthcare will be delivered, paid for, organized and this revolution will be driven by data and predictive analytics to make the most important use out of available historical data.



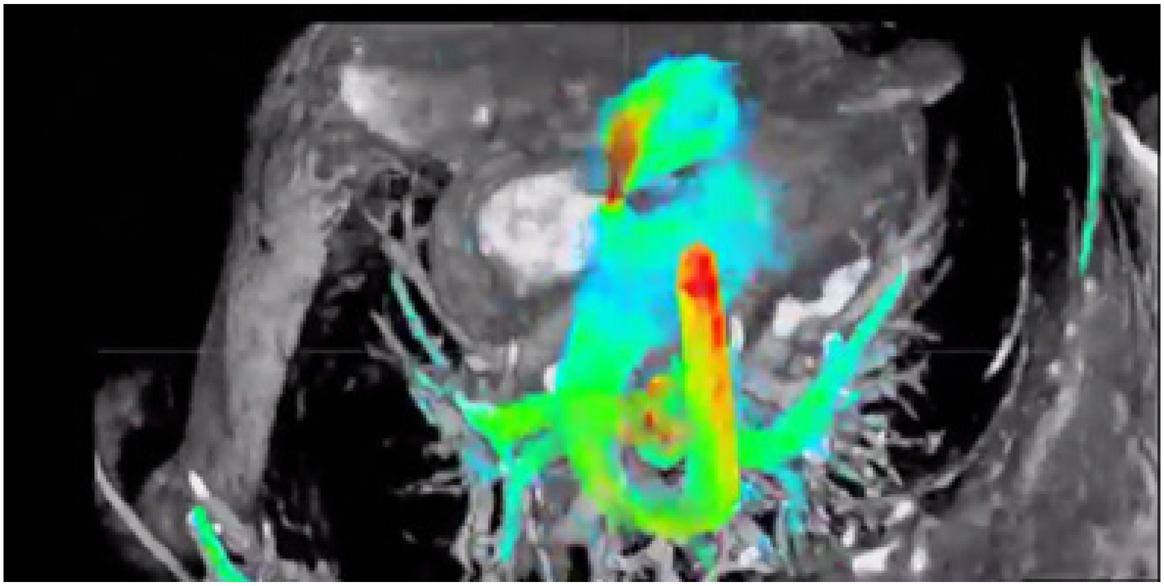
The Rise of Deep Learning

Healthcare and life sciences organizations are taking advantage of Artificial Intelligence (AI) and deep learning systems now that contemporary hardware and software architectures are able to keep pace with the demands of this class of machine learning. AI has been around for a long time, dating back to the 1950s, but waned in interest because compute and storage resources were not robust enough to address the needs of the technology. In the past few years, however, processing capabilities have caught up with these needs.

A group of researchers at Johns Hopkins University developed a novel approach applying deep neural networks (DNNs) to predict pharmacologic properties of many drugs. In this study, scientists trained DNNs to predict the therapeutic use of a large number of drugs using gene expression data obtained from high-throughput experiments on human cell lines. Authors used a sophisticated approach of measuring the differential signaling pathway activation score for a large number of pathways to reduce the dimensionality of the data while retaining biological relevance and used these scores to train the deep neural networks. The group also is developing multimodal DNNs to predict a broad range of properties of drugs, small molecules and natural compounds for a range of applications including treating common and rare diseases, aging, regenerative medicine and increasing response rates in cancer immunotherapy.

Additionally, Insilico Medicine, Inc. scientists published the first deep learned biomarker of human age aiming to predict the health status of the patient in a paper titled "Deep biomarkers of human aging: Application of deep neural networks to biomarker development" by Putin et al, in Aging and an overview of recent advances in deep learning in a paper titled "Applications of Deep Learning in Biomedicine" by Mamoshina et al, also in Molecular Pharmaceutics.

Illustrated in the figure below, deep learning medical imaging technology continues to advance improved imaging capabilities bringing artificial intelligence to the healthcare field, beginning with the heart. On yet another trajectory, some new applications leverage big data and high performance computing (HPC) to improve the lives of patients while increasing the efficiency and effectiveness of the healthcare system.



Distributed Systems – the Key to Success

Many organizations in the healthcare and life sciences industries are just now exploring the opportunities surrounding the dominant entries in the distributed processing architectures, Hadoop and Spark, and are looking for ways for getting started. Even though Hadoop has been in the market since 2011, it is still a relatively new technology in these sectors, while oftentimes Spark is for phase 2, for different use cases and for the advanced user.

Difficult challenges and choices face today's healthcare industry—researchers, clinicians and administrators have to make important decisions—often without sufficient data. Distributed systems like Hadoop and Spark offer open source platforms to make healthcare data available and actionable—researchers explore the genetic architecture of cancer cells; nurses and physicians monitor intensive care patients; administrators submit reimbursement claims before patients leave the hospital. Distributed computing systems are transforming healthcare.

Healthcare providers can get more valuable insights, manage costs, and provide better care options to patients by using data analytics. Big data technologies are enabling providers to store, analyze, and correlate various data sources to extrapolate knowledge. Benefits include efficient clinical decision support, lower administrative costs, faster fraud detection, and streamlined data exchange formats. It is projected that adoption of health data analytics will increase to almost 50 percent by 2017 from 10 percent in 2011, representing a 37.9 percent compound annual growth rate.

Hadoop is a strong example of a technology that allows healthcare to store data in its native form. If Hadoop didn't exist, decisions would have to be made about what can be incorporated into the data warehouse or the electronic medical record (and what cannot). Now everything can be brought into Hadoop, regardless of data format or speed of ingest. If a new data source is found, it can be stored immediately. No data is left behind. By the end of 2017, the number of health records of millions of people is likely to increase into tens of billions. Thus, the computing technology and infrastructure must be able to render a cost efficient implementation of:

(i) parallel data processing that is unconstrained
(ii) provide storage for billions and trillions of unstructured data sets
(iii) fault tolerance along with high availability of the system

Hadoop technology is successful in meeting the above challenges faced by the healthcare industry as the MapReduce engine and Hadoop Distributed File System (HDFS) have the capability to process thousands of terabytes of data. Hadoop makes use of highly optimized, yet inexpensive commodity hardware making it a budget friendly investment for the healthcare industry.

Hadoop Use Cases

In conventional IT environments, clinical, operational and financial data are managed in data silos. Meanwhile, with the movement from paper-based to electronic health records, and with the increase in usage of machines and medical devices that produce steady streams of data, the volume of data that healthcare institutions capture and analyze has skyrocketed, while the variety of that data has grown.

The Hadoop platform allows healthcare organizations to process and manage an ever-larger influx of data in a secure and cost-effective manner to improve quality and affordability. They can leverage the platform to bring together large volumes of detailed data from diverse sources, in a variety of formats, and consolidate it into a single flexible and scalable system for long-term storage and analysis.

- **Cancer treatments and genomics –** There has been an uptake in adopting Hadoop in the life sciences community, mostly targeting next-generation sequencing, and simple read mapping because what developers discovered was that a number of bioinformatics problems transferred very well to Hadoop, especially at scale.

www.insidebigdata.com | 508-259-8570 | Kevin@insideHPC.com

- **Monitoring patient vitals –** There are several hospitals across the world that use Hadoop to help the hospital staff work efficiently with big data. Without Hadoop, most patient care systems could not even imagine working with unstructured data for analysis.
- **Hospital networks –** Hadoop technology is used to help medical experts analyze high velocity data in real time from diverse sources such as financial data, payroll data, and EHRs.
- **Healthcare intelligence –** Hadoop technology is used to cultivate healthcare intelligence applications that assist hospitals, payers and healthcare agencies increase their competitive advantages by devising smart business solutions.
- **Fraud detection and prevention –** Using Hadoop technology, insurance companies have been successful in developing predictive models to identify fraud by making use of real-time and historical data of medical claims, weather data, wages, voice recordings, demographics, cost of attorneys and call center notes. Hadoop's capability to store large unstructured data sets in NoSQL databases and using MapReduce to analyze this data helps in the analysis and detection of patterns in the field of fraud detection.

Spark Use Cases

- **Precision medicine –** The promise of precision medicine is a far-reaching goal that will require sweeping changes to the ways physicians treat patients, health data is collected, and global collaborative research is performed. Precision medicine typically describes an approach for treating and preventing disease that takes into account a patient's individual variation in genes, lifestyle, and environment. Achieving this mission relies on the intersection of several technology innovations and a major restructuring of health data to focus on the genetic makeup of an individual. The healthcare ecosystem has chosen a variety of tools and techniques for working with big data, but one tool that comes up again and again in many of the new architectures is Spark. Spark is already known for being a major player in big data analysis, but it is additionally uniquely capable in advancing genomics algorithms given the complex nature of genomics research.

Spark is already known for being a major player in big data analysis, but it is additionally uniquely capable in advancing genomics algorithms given the complex nature of genomics research.

- **Genomics algorithms –** The transitioning of today's popular genomics algorithms to Spark is one path that researchers are taking to take advantage of the distributed processing capabilities of the cloud. Many of these are already being built on top of Spark. Although Spark provides many infrastructure advantages, Spark still speaks the languages that are popular with the research community. Languages like SparkR make for an easy transition into the cloud.
- **Computational neuroscience –** One example of a research project taking advantage of Spark is the Howard Hughes Brain Institute. The project's goal is to understand brain function by monitoring and interpreting the activity of large networks of neurons during behavior. An hour of brain imaging for a mouse can yield 50-100 gigabytes of data. The researchers developed a library of analytical tools called Thunder which is based on Spark using the Python API along with existing libraries for scientific computing and visualization. The core of Thunder is expressing different neuroscience analyses in the language of RDD operations. Many computations such as summary statistics, regression and clustering can be parallelized using MapReduce.

The Impact of IoT on Healthcare and Life Sciences

Patient-centered care is on the verge of a game-changing makeover as consumers embrace wearable devices, home monitoring tools, and mobile health apps at a staggering rate. According to a new report published by Allied Market Research titled, "World Internet of Things (IoT) Healthcare Market – Opportunities and Forecasts, 2014-2021", the global IoT healthcare market is expected to reach $136.8 billion by 2021, recording a Combined Annual Growth Rate (CAGR) of 12.5% between 2015 and 2021. Services and system/software segments collectively occupy a dominant share in the global IoT healthcare market and is expected to drive the growth over the forecast period. The patient monitoring application segment is expected to maintain its lead position with $72.7 billion by 2021.

The patient monitoring application segment is expected to maintain its lead position with $72.7 billion by 2021.

The global IoT healthcare market is anticipated to grow at a significant pace, owing to the increasing availability of wearable smart devices and decreasing cost of sensor technology. Furthermore, the launch of technologically advanced "smart" devices and analytics software, rising incidence rates of chronic diseases, surging demand for cost-effective treatment and disease management, better accessibility of high speed internet and implementation of favorable government regulatory policies, are also expected to fuel the growth of this market.

In addition, the improvement in healthcare infrastructure in developing economies, increase in government support, abundant R&D investments by major players for developing better IoT infrastructure are expected to offer potential growth opportunities to the market. Conversely, factors such as high costs associated with IoT infrastructure development, data privacy and security concerns, and limited technical expertise are projected to restrain the market growth.

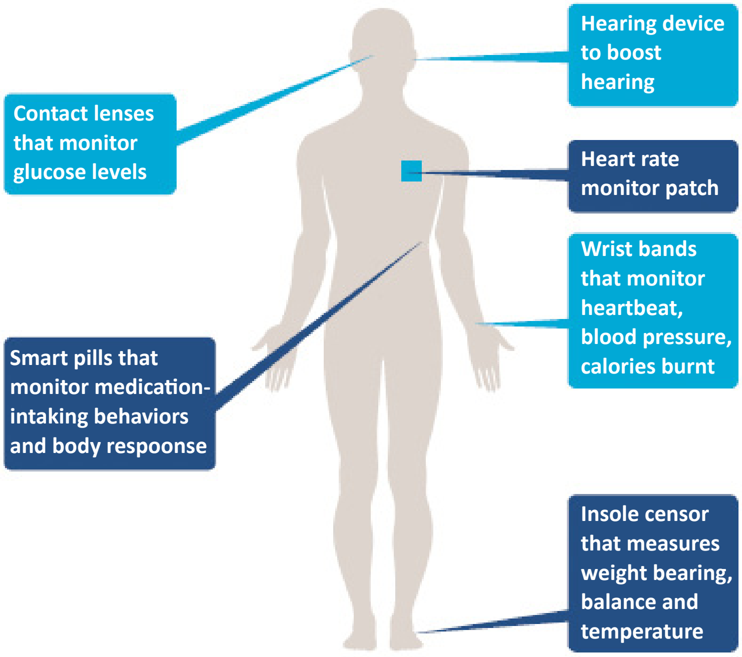
Despite numerous regulatory and privacy constraints, organizations inside and outside the healthcare industry are exploring ways to put the IoT to work. Players include pharma and biopharma manufacturers; hospitals and clinics; physicians, nurses and other healthcare providers; health insurers; fitness companies; and tech firms. The goals are to cut costs, boost efficiency and improve the way illnesses are diagnosed, treated and prevented.

At the same time, an increasing number of digitally empowered consumers are taking more responsibility for their health. Primed to use fitness wearables and smartphone apps, people are growing more comfortable with new types of sensors that capture and analyze their health and medical data. It will only be a matter of time before this information is seamlessly integrated into larger healthcare systems to make their care more precise and efficient.

Though their number is growing steadily, many IoT healthcare projects are still in their infancy, and remain a patchwork of disparate and isolated initiatives. And while it's not yet clear how things will shake out, there is no shortage of ideas. Many large and influential tech firms — including Alphabet, Apple, Dell EMC, General Electric, IBM, Intel, Microsoft, Philips, Samsung and SAP — have entered the IoT space in a big way and are hoping to make things happen quickly.

The result is that hospitals and healthcare systems are using the IoT to make their facilities more efficient. Initiatives include sharing records to ensure higher-quality care, tracking medical supply inventory and communicating with field personnel. Many pharma companies and medical device makers are already incorporating IoT components into their manufacturing and distribution operations. They are also exploring more strategic ways to harness it to make their products better during the research and development phase and in clinical trials.

The figure below provides compelling examples of how medical sensors might transform information and understanding of people's health status:



The Convergence of Big Data and HPC

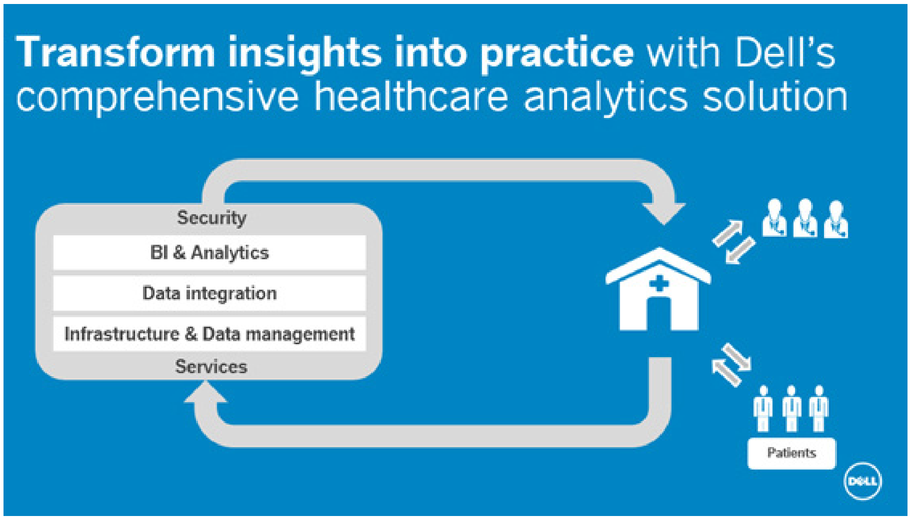
Another area that is of particular interest to the healthcare and life sciences fields is the increasing convergence of big data and high performance computing (HPC). As an example from the healthcare industry, researchers at John Hopkins University built a computational model of a typical heart using an HPC stack. This model simulates the function of the heart at multiple scales, from the molecular and cellular levels to structural and electrophysiological behavior, all with a great deal of accuracy. Here is where big data couples with HPC. With the convergence of big data technologies and HPC, we now have the opportunity to take terabyte-scale data sets from the MRI and CT scans of an individual patient and add them to the generic HPC heart model, allowing us to simulate the functioning of a specific patient's heart. The result is a powerful tool for planning narrowly targeted surgical procedures that drastically reduces the time required to perform each procedure and ultimately improving outcomes.

The compute nature of HPC is finding significant benefit from big data analytics and its ability to process high volume, high velocity data sets. The current most effective software platform for big data analytics — Hadoop — has its classic architecture consisting of HDFS and MapReduce, running on commodity cluster nodes, and the HPC environment has a different architecture where compute is distinct from the storage solution. You'd like to leverage your current investment in HPC by doing big data analytics on the architecture you already have. This is where two worlds come together.

There are many other powerful examples of how big data and HPC are converging in healthcare and life sciences, and that's good news for patients and the healthcare system. This combination will ultimately help clinicians save lives and improve patient outcomes while making the healthcare system more effective and more efficient.

Case Studies: Dell EMC Focused Customer Use Cases

This section highlights a number of high-profile case studies that are based on Dell EMC software and services and illustrate inroads into big data made by healthcare and life sciences organizations. Dell EMC provides an end-to-end portfolio of solutions and services for healthcare analytics, business intelligence and data management. Specifically, *Dell EMC Healthcare Services* and *Dell EMC BI & Analytics Services* maintain a strong focus on the healthcare domain and BI/analytics services and technology in order to fuel the transformation of healthcare with interconnected analytics solutions that drive quality improvements and cost reduction.



Healthcare Provider

Dell EMC performed a big data proof-of-value and data lake buildout for a healthcare provider. The goal was a big data solution for predictive analytics and enterprise reporting. The challenge was that the organization did not have an enterprise data warehouse to consolidate data from their enterprise operational systems. Further, key business stakeholders were not provided with tools needed to access data. Lastly, teams tasked with creating analytics spent 90% of their time integrating data in SAS, leaving them little time to analyze the data or create predictive analytics.

The successful project involved meeting with key business and IT stakeholders to determine reporting and analytic challenges and priorities, and also performing a Current State Assessment, along with meta data Discovery, profiling and outlier analysis of source data. Dell EMC proposed a data lake architecture to address enterprise reporting and predictive analytic needs. The solution also initiated a governance program to ensure data quality and to establish stewardship procedures. Finally, the project identified federated business data lake hardware and Pivotal big data suite software as the target platform for the data lake.

The results of the project included new client analytics environment that facilitated the execution of analytics and reporting activities to reduce time to insight. Further, client governance structure ensured that metadata for new data sources into the data lake was shared with users. The environment also supported the rapid creation of sandboxes to support analytics projects.

TGen

To advance health through genomic sequencing and personalized medicine, the *Translational Genomics Research Institute* (TGen) required a robust, scalable high-performance computing environment complimented with powerful big data analytics tools for its Dell EMC | Cloudera Apache Hadoop platform, accelerated by Intel. For example, a genome map is required before a personalized treatment for neuroblastoma can be designed for the patient, but conventional genome mapping takes up to 6 months. Dell EMC helped TGen reduce the time from 6 months to less than 4 hours and enables biopsy to treatment in less than 21 days. Now TGen is widely known for their comprehensive genome analysis and collaboration through efficient and innovative use of their technology to support researchers and clinicians to deliver on the promises of personalized medicine for better patient outcomes.

Arizona State University

ASU worked with Dell EMC to create a powerful HPC cluster that supports big data analytics. As a result, ASU built a holistic Next Generation Cyber Capability (NGCC) using Dell EMC and Intel technologies that is able to process structured and unstructured data, as well as support diverse biomedical genomics tools and platforms.

HPC technology and the Dell EMC | Cloudera Apache Hadoop solution, accelerated by Intel, upon which NGCC is based can handle data sets of more than 300 terabytes of genomic data. In addition, ASU is using the NGCC to understand certain types of cancer by analyzing patients' genetic sequences and mutations.

Transforming Patient Care at the University of Iowa Hospital and Clinics

Surgeons at the University of Iowa Hospitals and Clinics needed to know if patients were susceptible to infections in order to make critical treatment decisions in the operating room. The goals of the project were to predict patients with the biggest risk of surgical site infections, in order to reduce infection rate to improve patient care and decrease costs. Reducing the infection rate has major implications for overall patient health and cost savings. The solution consisted of the ability to merge historical EHR data and live patient vital signs to predict infection likelihood, and provide doctors with real-time, predictive decisions during surgical procedures so they can create a plan to reduce risk. The surgical team harnessed the power of big data analytics, coupled with other methods, to keep patients safe — reducing surgical site infections by 58 percent— while decreasing the cost of care and decreasing the incidence of readmissions by 40%.

> The solution consisted of the ability to merge historical EHR data and live patient vital signs to predict infection likelihood, and provide doctors with real-time, predictive decisions during surgical procedures so they can create a plan to reduce risk.

> Centegra Health System selected Dell EMC's Unified Clinical Archive (UCA) solution to manage its growing archive of diagnostic images. It gives unlimited scalability and includes an application-neutral onsite archive that replicates data to the cloud.

Centegra Health System

To boost patient care, avoid cyclical data migrations, achieve disaster recovery, and gain agility, Centegra sought a highly flexible and scalable solution for data management and archiving. Centegra Health System selected Dell EMC's Unified Clinical Archive (UCA) solution to manage its growing archive of diagnostic images. It gives unlimited scalability and includes an application-neutral onsite archive that replicates data to the cloud. The benefits of the solution include:

(i) boosting patient care, service levels and efficiency by simplifying data access
(ii) staff can view patient information with high reliability
(iii) saving money by avoiding data migrations and upgrades
(iv) increasing agility by breaking free of proprietary file constraints.

Large Healthcare Organization

A large healthcare organization was looking to improve their data warehouse response and capabilities with a proven MPP based appliance. Additionally, they need a sandbox for Hadoop in order to prototype unstructured workloads. The Dell EMC QuickStart for Cloudera Hadoop, accelerated by Intel was chosen to provide an engineered solution for their Hadoop testbed. The native software in Microsoft APS, called Polybase, was used to incorporate both structured (SQL) and unstructured (Hadoop) data using SQL skills for their Value Based Care business. The goal was to prove out some test cases and then roll out to other business units.

Summary

Big data technologies have the potential to transform the way healthcare and life science organizations use sophisticated technologies to gain insights from their clinical and other data repositories to make informed decisions. Big data enabled analytics allow organizations to investigate and explore data to identify trends, patterns, and relationships to reveal insights that, when combined with business context, serve to create knowledge. Many organizations are using big data analytics and having huge success. This is not a new concept. The issue becomes how to make the data work for your organization so it's the right data for the right patient at the right time.

> The healthcare industry is moving from simply reporting facts to discovery of insights with an emphasis toward empowering data-driven healthcare organizations.

Big data holds far reaching potential to change the entire healthcare value chain, from patient care personalization, to industrialization of healthcare provider processes for improved clinical outcomes and increased efficiency, to drug discovery, to safer public health management, to more inclusive and effective reimbursement of care. The healthcare industry is moving from simply reporting facts to discovery of insights with an emphasis toward empowering data-driven healthcare organizations. The aim is to turn information into actionable insights to prevent inefficiencies and adapt workflows for improved healthcare outcomes across the end-to-end patient journey.

The potential for big data solutions and the use of IoT in healthcare, however, is still generally untapped. It is not just a matter of semantics and data interoperability, it is more holistically a matter of understanding what set of skills, methodologies, organizational, and regulatory changes are necessary to leverage the benefits of big data. Realizing this opportunity will require an end-to-end strategy where IT is the technical enabler but where new organizational and process aspects are led by enterprise decision makers that will also set the overall business objectives. Such a comprehensive strategy needs to be developed through a step-by-step approach to be integrated into the healthcare organization information strategy as soon as possible.

Looking at future applications of big data in healthcare it is interesting to see how they further enhance and accelerate the convergence between the activities of clinicians, administrators, policy makers, payers, and researchers by saving costs, creating greater efficiencies based on outcome comparison, reducing risks, and improving personalized care. Patients can benefit from this convergence too because, besides being the ultimate sources of healthcare data, they can make more informed decisions about their health, playing a much more proactive role in their care paths.

> Relying on data lake technology from Dell EMC, the Integrated Data Environment for Analytics (IDEA) platform combines public domain data, institutional data as well as data from investigators.

A good example of what the future might hold is the Partners HealthCare system mentioned earlier — Integrated Data Environment for Analytics (IDEA) platform. Relying on data lake technology from Dell EMC, the platform combines public domain data, institutional data as well as data from investigators. In the future, the platform will be able to ingest imaging data from radiology and pathology sources, which will help to cultivate a rich and robust common data set that researchers may draw upon for their work. Data lakes are becoming a popular way to store massive volumes of healthcare data that may or may not have a distinct use case at the moment. This flexible class of data set provides researchers with the opportunity to query the database in ways they may not even have envisioned when the

Combined with the growing interest in application-based healthcare technologies and plug-and-play data standards, data lakes can become an important resource for vendors, developers, and academics looking to meet the challenging needs of providers and patients.

data was initially collected. Combined with the growing interest in application-based healthcare technologies and plug-and-play data standards, data lakes can become an important resource for vendors, developers, and academics looking to meet the challenging needs of providers and patients.

By all accounts, the implementation and use of big data technologies will continue to spread rapidly. To that end, several challenges must be addressed. Unlocking the value of data requires planning, organizational and physician buy-in, and highly skilled IT staff. The right data can reduce the cost of care by close to 15%, according to McKinsey. Additionally, as big data capabilities continue to evolve, concerns such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies need to be resolved in an ethical, compliant and cost-effective manner. Only then will organizations garner the true benefits from big data and analytics in healthcare and life sciences.