Uses of Big Data and Analytics In Healthcare

Arun Kumar Arumugam
Senior IT Architect,
IQVIA, Plymouth Meeting, PA, USA

Abstract:- **The paper describes the promise and potential of big data analytics in healthcare and the nascent field of big data analytics in healthcare, discusses the benefits, outlines an architectural framework and methodology, describes examples reported in the literature, briefly discusses the challenges, and offers conclusions. The paper provides a broad overview of big data analytics for healthcare researchers and practitioners. Big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. Its potential is great; however there remain challenges to overcome.**

Keywords:- Big Data, Healthcare, Big Data Analytics, Challenges in Healthcare.

I. INTRODUCTION

Healthcare industry is growing in larger pace and hence the Healthcare data from different source. Due to larger volume of data, it is hard to process the data and avoid data loss. Big Data can be used to resolve the data processing issues and run more analytics on data which could be used for lot of purposes. It also helps us to fetch relevant information from data warehouse and effectively use the data to provide treatment on time and prescribe medicines. Let's discuss more in detail.

II. DATA PROBLEM IN HEALTH CARE INDUSTRY

Data Collection: A Messy Problem
Each year, over 1.2 billion clinical documents are produced in the U.S., and this number is growing at a rate of 48% per year. That's a ton of data. Unfortunately, 80% of that healthcare data is sitting around in unstructured formats and locked away in myriad Electronic Health Record Systems (EHRs).

Electronic health records (EHRs) are used to capture and manage information collected during patient appointments. Other data sources for patient profiles include personal health records (PHRs) and patient portals, claims and reimbursement information from payers. The health information exchange (HIE) is yet another initiative expediting the movement and consolidation of data between various care partners. Apart from this, many other variables may contribute to health outcomes.

Below are the four major problems faced by Healthcare:
1. *Fragmented Data*
   a. Health care data comes from a bewildering number of sources and different formats, such as structured data, paper, digital, pictures, videos, multimedia and so on. Data collection and aggregation communities are equally fragmented,

making the extraction and integration of data a real challenge.
   b. Providers, payers, public health specialists, employers, social network communities and patients all collect data, but there is no effort to unify the information
2. *Ever-changing Data*
   a. Patients and physicians, like everyone else, move, change their names and professions, retire and die. Payer organizations may also relocate, add new locations or go through various mergers and acquisitions
   b. Stale data and information latency directly impact a member's experience and providers' business sustainability. The result is a delay in the adoption of new treatment options, inadequate response to health care programs and poor engagement and experience.
3. *Privacy and Security Regulations*
   a. Maintaining patient trust is the cornerstone for building an efficient health care ecosystem. Data security has become of utmost importance to the health care industry as patient privacy depends on HIPAA2 compliance and secure adoption of electronic health records
   b. Poor data quality and strategy prevent organizations from meeting new regulatory needs and result in high costs associated with audits and reporting.
4. *Patient Expectations*
   a. The health care industry is about to experience the same shift we saw in retail, banking and hospitality. The health care system is on the verge of a perfect storm. The industry will need to understand members' changing needs and their preferences and then provide solutions that align with their way of life.

III. WHAT IS BIG DATA?

Big data refers to the large, diverse sets of information that grow at ever-increasing rates. It encompasses the volume of information, the velocity or speed at which it is created and collected, and the variety or scope of the data points being covered. Big data often comes from multiple sources like Oracle, Postgres, Hadoop, Mango, etc., and arrives in multiple formats like JSON, string, CSV, etc.,

IV. HOW BIG DATA WORKS?

Big data can be categorized as unstructured or structured. Structured data consists of information already managed by the organization in databases and spreadsheets; it is frequently

numeric in nature. Unstructured data is information that is unorganized and does not fall into a pre-determined model or format. Big data is most often stored in computer databases and is analyzed using software specifically designed to handle large, complex data sets. Many software-as-a-service (SaaS) companies specialize in managing this type of complex data.

Big Data helps us to organize and format the unstructured data from various sources. It helps to run analytics on the data to get relevant information about the patient without any data loss and synchronize data on a timely basis. The analytics is used for various other purposes like diagnosis of a new disease from the vitals noted and so on. Below image depicts how big data works by converting unstructured data and analytics.
*
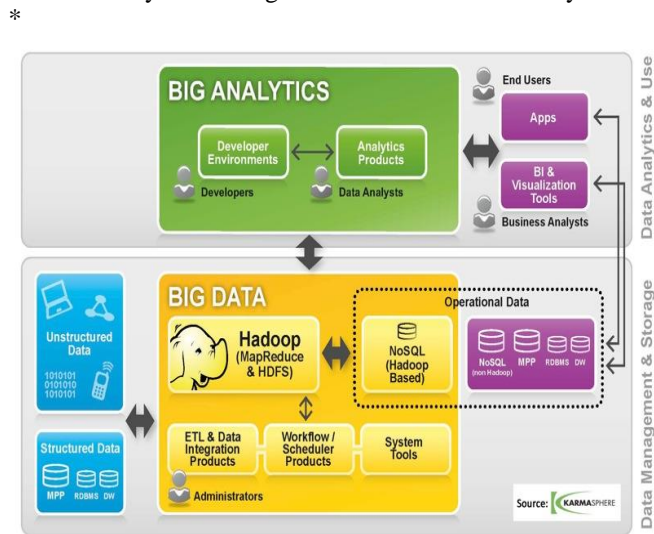
V.   3V'S OF BIG DATA
Big Data can be defined using the famous 3 Vs - Volume, Velocity and Variety.

1. **VOLUME** - Volume refers to the amount of data generated through websites, portals and online application

2. **VELOCITY –** Velocity refers to the speed with which data is being generated.

3. **VARIETY –** Variety refers to all the structured and unstructured data that has the possibility of getting generated either by humans or by machines.

VI.   TYPES OF ANALYTICS ON BIG DATA IN HEALTHCARE
Big Data has huge potential to change the Medical Industry; big data tools gather billions of data points that can be utilized for health management in three main areas:
1. **Descriptive analytics**: It calculates what has happened, such as cost, frequency and resources.
2. **Predictive analytics:** It uses the descriptive data to predicate expected results in the future.

3. **Prescriptive analytics**: It offers the ability to make positive decisions considering anticipating predictions.
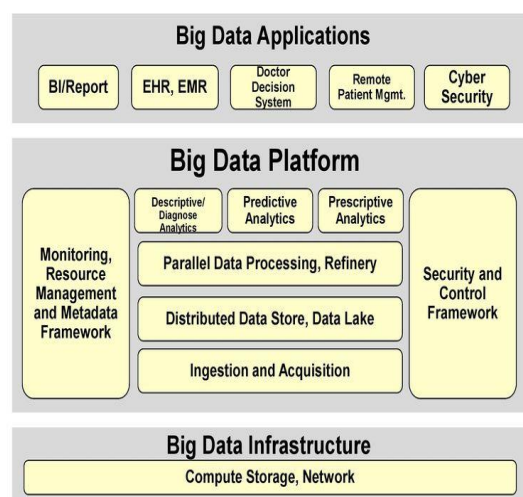
VII.   BIG DATA IN HEALTCARE
"Big data in healthcare" refers to the abundant health data amassed from numerous sources including electronic health records (EHRs), medical imaging, genomic sequencing, payor records, pharmaceutical research, wearables, and medical devices, to name a few.

Three characteristics distinguish it from traditional electronic medical and human health data used for decision-making: It is available in extraordinarily high volume; it moves at high velocity and spans the health industry's massive digital universe; and, because it derives from many sources, it is highly variable in structure and nature. Below are the various sources of data for healthcare:

1. Government Agencies
2. Patient Portals
3. Research Studies
4. Payer Records
5. Generic Database
6. Smart Phone
7. Wearable Technologies
8. Search engine data
9. Public record data
10. Electronic Health Records (HER) and so on

With its diversity in format, type, and context, it is difficult to merge big healthcare data into conventional databases, making it enormously challenging to process, and hard for industry leaders to harness its significant promise to transform the industry.

Big Data Architecture for Healthcare

Big Data helps healthcare industry to convert data from various sources to a useful, actionable information. By leveraging appropriate software tools, big data is informing the

movement toward value-based healthcare and is opening the door to remarkable advancements, even while reducing costs. With the wealth of information that healthcare data analytics provides, caregivers and administrators can now make better medical and financial decisions while still delivering an ever-increasing quality of patient care.

There are at least two trends today that encourage the healthcare industry to embrace big data.

1. A pay-for-service model, which financially rewards caregivers for performing procedures, to a value-based care model, which rewards them based on the health of their patient populations. Healthcare data analytics will enable the measurement and tracking of population health, thereby enabling this switch.

2. Using big data analysis to deliver information that is evidence-based and will, over time, increase efficiencies and help sharpen our understanding of the best practices associated with any disease, injury or illness.

Undoubtedly, adopting the use of healthcare big data can transform the industry, driving it away from a fee-for-service model toward value-based care. In short, it can deliver on the promise of lowering healthcare costs while revealing ways to deliver superior patient experiences, treatments, and outcomes.

VIII. APPLICATIONS OF BIG DATA IN HEALTHCARE

According to Healthcare industry, keeping patients healthy and avoiding illness and disease stands at the front of priority list. The resulting data is already being sent to cloud servers, providing information to physicians who use it as part of their overall health and wellness programs.

Below are the ways to apply Big Data in Healthcare

Diagnostics	Data mining and analysis to identify causes of illness.
Preventative medicine	Predictive analytics and data analysis of generic, lifestyle and social circumstance to prevent disease
Precision Medicine	Leveraging aggregate data to drive hyper-personalized care.
Precision Medicine	Data driven medical and pharmacological research to cure diseases and discover new treatments and medicines
Reduction of adverse medication events	Harnessing of big data to spot medication errors and flag potential adverse reactions
Cost reduction	Identification of value that drives better patient outcome for long term savings
Population Health	Monitoring Big data to identify disease trends and health strategies based on demographics, geography and socio-economics

IX. CHALLENGES FOR IMPLEMENTING BIG DATA IN HEALTHCARE

Healthcare organizations face challenges with healthcare data that fall into several major categories including data

IX. MANAGEMENT AND ANALYSIS OF BIG DATA

Big data is the huge amounts of a variety of data generated at a rapid rate. The data gathered from various sources is mostly required for optimizing consumer services rather than consumer consumption. This is also true for big data from the biomedical research and healthcare. The major challenge with big data is how to handle this large volume of information. To make it available for scientific community, the data is required to be stored in a file format that is easily accessible and readable for an efficient analysis. In the context of healthcare data, another major challenge is the implementation of high-end computing tools, protocols and high-end hardware in the clinical setting. Experts from diverse backgrounds including biology, information technology, statistics, and mathematics are required to work together to achieve this goal. The data collected using the sensors can be made available on a storage cloud with pre-installed software tools developed by analytic tool developers. These tools would have data mining and ML functions developed by AI experts to convert the information stored as data into knowledge. Upon implementation, it would enhance the efficiency of acquiring, storing, analyzing, and visualization of big data from healthcare. The main task is to annotate, integrate, and present this complex data in an appropriate manner for a better understanding. In absence of such relevant information, the (healthcare) data remains quite cloudy and may not lead the biomedical researchers any further. Finally, visualization tools developed by computer graphics designers can efficiently display this newly gained knowledge.

Heterogeneity of data is another challenge in big data analysis. The huge size and highly heterogeneous nature of big data in healthcare renders it relatively less informative using the conventional technologies. The most common platforms for operating the software framework that assists big data analysis are high power computing clusters accessed via grid computing infrastructures. Cloud computing is such a system that has virtualized storage technologies and provides reliable services. It offers high reliability, scalability and autonomy along with ubiquitous access, dynamic resource discovery and composability. Such platforms can act as a receiver of data from the ubiquitous sensors, as a computer to analyze and interpret the data, as well as providing the user with easy to understand web-based visualization. In IoT, the big data processing and analytics can be performed closer to data source using the services of mobile edge computing cloudlets and fog computing. Advanced algorithms are required to implement ML and AI approaches for big data analysis on computing clusters. A programming language suitable for working on big data (e.g. Python, R or other languages) could be used to write such algorithms or software. Therefore, a good knowledge of biology and IT is required to handle the big data from biomedical research. Such a combination of both the trades usually fits for bioinformaticians. The most common among various platforms used for working with big data include Hadoop and Apache Spark.

aggregation, policy and process, and management. Let's explore these further.

Data Aggregation Challenges. First, patient and financial data are often spread across many payors, hospitals,

administrative offices, government agencies, servers and file cabinets. Pulling it together and arranging for all data producers to collaborate in the future as new data is produced requires a lot of planning. In addition, every participating organization must understand and agree upon the types and formats of big data they intend to analyze.

Management Challenges. Finally, realizing the promises of big data analytics in healthcare requires organizations to adjust their ways of doing business. Data scientists will likely be needed along with IT staff that have the required skills to run the analytics. Some organizations may struggle with the thought of having to "rip and replace"

Policy and Process Challenges. Once data is validated and aggregated, various process- and policy-related issues need to be addressed. The HIPAA regulations demand that policies and procedures protect health information. Access control, authentication, security during transmission, and other rules complicate the task. This multifaceted issue has been solved to some extent by cloud service providers, perhaps most notably Amazon AWS, which offers cloud services that comply with HIPAA and Protected Health Information (PHI).

X. THE BRILLIANT FUTURE FOR BIG DATA IN HEALTHCARE

Industrial sectors declare their big data initiatives have been successful and transformational, the outlook for healthcare is even more exciting. Below are a few areas where big data is destined to transform healthcare.

Precision medicine, as envisioned by the National Institutes of Health, seeks to enroll one million people to volunteer their health information in the All of Us research program. That program is part of the NIH Precision Medicine Initiative. According to the NIH, the initiative intends to "understand how a person's genetics, environment, and lifestyle can help determine the best approach to prevent or treat disease. The long-term goals of the Precision Medicine Initiative focus on bringing precision medicine to all areas of health and healthcare on a large scale."

Wearables and IoT sensors, already noted above, have the potential to revolutionize healthcare for many patient populations and to help people remain healthy. A wearable device or sensor may one day provide a direct, real-time feed to a patient's electronic health records, which allows medical staff to monitor and then consult with the patient, either face-to-face or remotely.

Machine learning, a component of artificial intelligence, and one that depends on big data is already helping physicians improve patient care. IBM with its Watson Health computer system has already partnered with Mayo Clinic, CVS Health, Memorial Sloan Kettering Cancer Center, and others.

CONCLUSION

In this article, various uses and challenges of implementing Big Data in healthcare have been discussed. Big Data can be effectively used to maintain the growing volume of healthcare data as the healthcare data is larger volume with different data formats. This data can be used in healthcare for various purposes like analyzing the patient vitals, diagnosis and alert patients about the various health related problems.

REFERENCES

[1] A Practitioner's Guide to Business Analytics: Using Data Analysis Tools to Improve Your Organization's Decision Making and Strategy, R. Bartlett
[2] Big Data: Principles and Best Practices of Scalable Real-Time Data Systems By Nathan Marz And James Warren
[3] Big Data in Practice By Bernard Marr
[4] References from internet
[5] Analytics In A Big Data World: The Essential Guide To Data Science And Its Applications By Bart Baesens