

# Big Data – Next Technology for Healthcare

Rachana Rajendra Patil<sup>1</sup>

<sup>1</sup>Vidyalankar Institute of Technology, Mumbai, India

**Abstract:** *Big Data is next step technology towards a more resourceful and more informative knowledge and analysis of many different aspects which are hidden as it provide analysis based on both structured and unstructured. Here we have discussed on the healthcare system which gets benefited by Big Data.*

**Keywords:** Big Data, Healthcare, Cloud, HADOOP, SaaS, Teradata.

## 1. Introduction

This is an age of open information in healthcare industry. During the last decade, we have seen a huge progress in converting medical records to digital form where several pharmaceutical companies and other organizations consolidated years of R&D data and stored them in electronic databases. This has resulted in accumulation of an enormous amount of data through these years in various sector of industries. Now, the need of the hour is to organize this heap of data, make them usable, searchable and actionable. And here comes the picture of Big Data in a broader way.

Big data creates new problem solving approaches by surpassing the abilities of conventional storage, diagnostic and reporting systems. With the continuous advent of more advanced technologies in the arena of database storage techniques, computing, wireless data, mobile & social networking, processing big data is now possible in a number of profitable ways. And likely the way business decisions were being made till now, are getting changed by big data.

Big data solutions primarily target cost-effective solutions for the challenges of huge and continuously growing data and capture the potential analytical value it has. An effective and optimum analysis of this huge amount data would enable the user in finding the root cause and perform predictive analysis that enables understanding of what is likely to happen in future and preventive or corrective measures could be taken accordingly. For example, the common healthcare stakeholders, like members, providers, groups, payers could have a direct impact by big data which can predict their likely behavior by effectively analyzing their historical records and determine their desirable and undesirable behaviors.

The purpose of this paper is to define big data, have a deeper look into the opportunities and challenges it poses for the healthcare industry and discuss technology solutions.

## 2. What is Big Data?

Big data could be defined as a large amount of data which meets three criteria: Volume, Variety and Velocity.

- **Volume:** It's concerned with the amount of data today organizations are dealing with. It changed from GBs to Tera/Peta/Exa/Zetta bytes. This huge amount of data requires ascendable storage facility which can support

composite query mechanisms across several data sources.

- **Variety:** The data may be structured, unstructured or semi structured. Previous technologies used to convert unstructured data into structured form to analyze. But Big Data can handle all these three types of data. While standard methods and technologies occur to deal with huge volumes of structured data, it becomes a substantial challenge to analyze and process a large amount of highly inconstant data and turn it into actionable information. But this is where the huge potential of big data lays.
- **Velocity:** It's concerned with how fast we grab data and how fast we update existing data. Big data is developed to handle large volume of data streaming in with high rates. While traditional data warehouse analytics incline to be centered on periodic — daily, weekly or monthly — loads and updates of data, big data is treated and evaluated in real- or near-real-time. This is significant in healthcare for areas such as clinical decision support, where access to up-to-date information is crucial for precise and timely decision-making and omission of errors.

## 3. Big Data Leads to Big Opportunities in Healthcare

Big data has an important part in the future analytics environment which can influence patients, providers, researchers, payers and other healthcare constituents in various ways. This will affect how these components get absorbed with the existing healthcare network, especially when external information, eHealth and mHealth are involved.

Big data is leading to a transposal in the healthcare system. In the age old model, patients were kept in long term treatments by facilities and other providers as more inpatient days in turn brought more revenue. The new trend, including ACOs, encourages providers to keep patients healthy rather than in hospitals.

At the same time, patients are the most important element of Health Care industry. They are becoming increasingly demanding seeking more and more information to make their choice wisely. So, providing them with proper and latest evidence and information is very important that would facilitate them in making the best choice and aligning themselves to the latest treatments available.

In addition to demographics and medical history, the information that patients divulge about themselves are very beneficial reducing costs and improve treatment. Following are some of the challenges that exist with self-reported data:

- **Accuracy:** Generally, people tend to understate their negative points such as smoking or drinking habits while overstate the positive ones such as healthy diet, exercise. These erroneous data can be corrected through Big Data which would improve accuracy.
- **Privacy concerns:** People are reluctant to reveal their history because of privacy and other concerns. Effective methodologies and reassurances must be placed to guarantee the privacy of the data that patients submit from external access.
- **Consistency:** Certain yard sticks need to be established and applied to encourage uniformity in self-reported data across the healthcare system to eliminate local differences and increase the effectiveness of data.
- **Facility:** Methodologies such as mobility and social networking based on eHealth and mHealth need to be productively incorporated which would ease members' capability to self-report.

#### 4. Big Data Challenges

The present problem in healthcare industry is not the dearth of data but the dearth of information which could be used for precise and accurate decision making, plan and chalking out strategies.

As an instance, hospitalization of a single patient produces a number of data components including diagnoses, medicines, nursing history, lab tests, billing etc. These data belonging to a single patient needs to be analyzed, processed and consolidated into a huge data source to

When this this is multiplied by all the patient hospitalizations across the system and combined with multiple points where the data is produced and stored, the scope of the big data challenge emerges from there.

Below are some of the specific challenges of healthcare big data:

- **Fragmentation** - The prime hurdle to effective use of big data is the nature of healthcare information. Payers, providers, research centers and other constituents all have their own form of data. These are primarily difficult to assimilate because of concerns about confidentiality and propriety, composite and fragmented nature of the data within each silo. Even if everyone shared their data, there would be adequate challenges assimilating it within the silo.
- **Security** - The entire healthcare structure can reap profits from democratizing big data access. The cloud makes divulging and sharing big data easy and relatively inexpensive. However, significant security and privacy concerns exist, including the Health Insurance Portability and Accountability Act (HIPAA). A proper authentication process could ease and automate this access, but there are difficulties and challenges. Data access should be

controlled by group, role and function. Lastly, the security of the data once it leaves the cloud also needs to be guaranteed.

- **Timeliness** - Clinical decision support, which is one of the most crucial part of the benefits that Big Data has to offer to healthcare industry, depends wholly on the timeliness of the data it requires for making precise and accurate decisions.

In some situations, there is a very limited window for clinical decision support significantly smaller than the time it takes to run a report or analytic query. As big data arrive to the decisions based on a huge volume of latest and relevant data, which makes the decision making process simpler, faster and more accurate.

#### 5. Technology Choices for Big Data Solution

Technology solutions for dealing with big data offers a spectrum of choices to the user. There are on-site, cloud, open source or proprietary a vast range of technology choices are currently active in the market. For example, on site choices include Teradata, Vertica (HP) and Netezza (IBM). Whereas these solutions offers low time to value and maintenance but they are quite relatively high.

- **Hadoop:** This is an open Source framework used by various organizations as a productive, accessible and cheap option for dealing with big data. Proper training, proficient services and skilled support are needed to effectively deploy Hadoop solutions using the open source framework. Vendors such as Greenplum (a division of EMC), Microsoft, IBM and Oracle have commercialized Hadoop and united it with the rest of their database and systematic assistances.
- **Cloud Hosted Software as a service (SaaS):** Many renowned organizations which deals with huge amount of data have implemented the MapReduce-based solutions. It processes a huge amount of data belonging to a large number of computers. This algorithm divide large problems into a set of distinct tasks which is distributed to a large number of computers for getting the data processed and obtain a resultant solution. Tableau is another cloud based solution that supports visualization.
- **SaaS** is another significant technology for democratizing the results of big data. It is sometimes referred to as "on-demand software". SaaS is typically accessed by users using a thin client via a web browser. The vast majority of SaaS solutions are based on a multi-tenant architecture. With this model, a single version of the application, with a single configuration (hardware, network, operating system), is used for all customers ("tenants"). To support scalability, the application is installed on multiple machines (called horizontal scaling).

#### 6. Recommendations

To successfully identify and implement big data solutions and benefit from the value that big data can bring, healthcare organizations need to devote time and resources to visioning and planning. This will provide the foundation needed for

strong execution. Without this preparation, organizations will not realize the envisioned benefits of big data and will risk being left behind competitors.

## References

[1] Big data: The Next Frontier for innovation, Competition and Productivity – McKinsey Global Institute Report.

## Author Profile

**Rachana Rajendra Patil** pursuing M. E. degree in Computer Science from Vidyalankar Institute of Technology, Mumbai in year 2013-2015 and received the B.E. degrees in Computer Engineering from Shivajirao S. Jondhale College of Engineering in year 2010 - 2013. Very passionate about learning new technology and creating and implementing ideas for making complicate things simpler and intresting.

