

# A Proposed Framework to Enhance Healthcare Systems Utilizing Data Analytics for Intelligent Decision Support Systems

**Asaad El-Saadany Beder**

PhD. Student  
(SC&MIS Department)  
Asaad\_1231@yahoo.com

**Mohamed M. EL HADI**

Professor at Computer and  
Information Systems Depart-  
ment, Sadat Academy for  
Management Science,  
Cairo, Egypt

## Abstract

The healthcare system has generated large amount of data (Big Data) generated from electronic record. To improve the quality of healthcare, it's necessary that large volume of data generated should be analyzed effectively to answer new challenges. But dealing with this huge amount of data required Analytical embrace ideas that go beyond the Traditional statistical methods already in use in the Egyptian context. Big data analytics helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning algorithms, Data Mining and Artificial Intelligence. It requires a technology that helps to perform a real time analysis on the enormous data set, which is not utilized in the Egyptian health care information system. Modern analytics of the Egyptian health care system is still missing to help national decision making. Utilizing advanced data analytics for improving Egyptian health care policies and decision making. This research proposed big data analytics framework applied on the healthcare system to get the greatest value from their data and information by using hadoop map/reduce framework.

Keywords: Analytics, big data, big data analytics, Health Informatics, decision support system, and hadoop map/reduce.

## 1. Introduction

Health Informatics is one of the rapidly growing fields that focus on applying "Computer Science" and "Information Technology" to medical and health data. The healthcare system historically has generated large amounts of data, driven by record keeping, regulatory requirements, and patient care. While most data are stored in hard copy form, the current trend is toward rapid digitization of these large amounts of data. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data (known as 'big data') hold the promise of supporting a wide range of medical and healthcare functions, to derive previously untapped intelligence and insights from data to address many new and important questions. (Wullianallur and Viju, 2014).

Big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods.

By discovering associations and understanding patterns and trends within the data, big data analytics has the potential to improve care, save lives and lower costs. Thus, big data analytics in healthcare take advantage of the explosion in data to extract insights for making better informed decisions.

Analytics when applied in the context of big data is the process of examining large amounts of data, from a variety of data sources and in different formats, to deliver insights that can

enable decisions in real or near real time. Various analytical concepts such as data mining, natural language processing, artificial intelligence and predictive analytics can be employed to analyze, contextualize and visualize the data. Big data analytical approaches can be employed to recognize inherent patterns, correlations and anomalies which can be discovered as a result of integrating vast amounts of data from different data sets (H.V. Jagadish et al., 2012).

The researcher conducted an exploratory study included Dakahlia Health Directorate and the Center of Information has been done through a number of personal interviews with some of the workers.

The Ministry of Health and Population, representing the executive branch and is responsible for providing healthcare for citizens, through a vast network of health service units.

Data will flow from the health units of the Ministry of Health and Population as follow:

- There is only one department responsible for data quality and revision on the lower levels associated with data activities. Using the data collected in the calculation of health indicators as a management tool.
- The information produced by health service units, which are many and far reaching. The health clerk record and monitor the number of cases in the daily record. And send weekly, monthly and annually disclosure to the center of Health Information. In the case of large units such as hospitals their statistics office in different sections within the health service unit (outpatient - inpatient - labs).
- The Information center compiled the data and enters the data to the computer. The Information Center sends data files to the National Center for Health Information.
- Available data are curative care data to public and central hospitals (public hospital services), primary healthcare for Motherhood and Childhood Data

## 2. Analytics

A starting point for understanding analytics is to explore its roots. Decision support systems (DSS) in the 1970s were the first systems to support decision making. DSS came to be used as a description for an application and an academic discipline. Over time, additional decision support applications such as executive information systems, online analytical processing (OLAP). Then in the 1990s, Howard Dresner, an analyst at Gartner, popularized the term business intelligence. A typical definition is that "BI is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions". With this definition, BI can be viewed as an umbrella term for

all applications that support decision making, and this is how it is interpreted in industry and, increasingly, in academia. BI evolved from DSS, and analytics evolved from BI (at least in terms of terminology). Thus, analytics is an umbrella term for data analysis applications.

BI can also be viewed as "getting data in" (to a data mart or warehouse) and "getting data out" (analyzing the data that is stored). A second interpretation of analytics is that it is the "getting data out" part of BI. The third interpretation is that analytics is the use of "rocket science" algorithms (e.g., machine learning, neural networks) to analyze data.

These different takes on analytics do not normally cause much confusion, because the context usually makes the meaning clear. The progression from DSS to BI to analytics is shown in Figure 1.

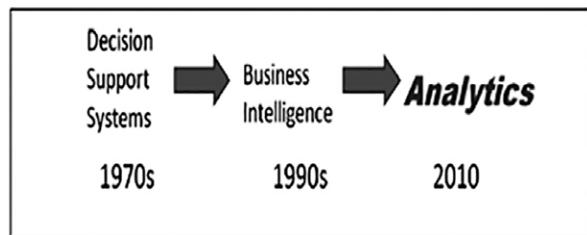
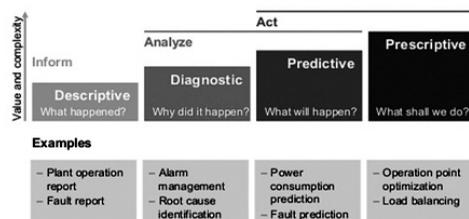


Figure 1: from DSS to BI to analytics

Adams J and Klein J, (2011).

3. Data Analytics: Guide decision making by discovering patterns in data using statistics, programming, and operations research (Michael Minelli, 2013)

- Descriptive Analytics: Analyzing historical data to explain past success or failures.
- Predictive Analytics: Forecasting using historical data.
- Prescriptive Analytics: This type of analysis reveals what actions should be taken. This is the most valuable kind of analysis and usually results in rules and recommendations for next steps.
- Diagnostic Analytics - A look at past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.



Gartner 2013 - N. Gauss/Siemens - 2015

Carvia - N. Boujmaa - System Numérique - Matière Big Data - 16 Juin 2016

Figure 2: Types of Analytics

(Michael Minelli, 2013)

#### 4. Big Data Lifecycle

- **Data Collection:** It involves the collection of data from various sources and storing it in HDFS. Data can be anything such as case history, medical images, social logs, sensor data etc.
- **Data Cleaning:** It involves the process of verifying whether there is any junk data or any data that has missed values. Such data needs to be removed.
- **Data Classification:** It involves the filtering of data based on their structure. For example Medical Big data consists of mostly unstructured data such as hand written physician notes. Structured, semi-structured and unstructured data should be classified in order to perform meaningful analysis.
- **Data Modeling:** It involves performing analysis on the classified data.
- **Data Delivery:** It involves the generation of report based on the data modeling done.

### 5. Big Data Analytics in Healthcare

For a big data analytics project in healthcare, the conceptual framework is similar to that of a traditional health informatics or analytics project. But, how processing is executed is totally different in both. The analysis can be performed with a business intelligence tool installed on a stand-alone system, like a laptop or a desktop in a regular health analytics project. Because according to the definition, big data is nothing but large processing distributed across the multiple nodes. This concept has existed for many years. To gain insight for making better-informed health-related decisions, as healthcare providers start to tap into their large data repositories, relatively new thing is its use in analyzing very large data sets. (Russom, P., 2011). Also, the applications of big data analytics in healthcare have been encouraged by open source platforms such as Hadoop/MapReduce, available on the cloud. The typical end-user in healthcare may not possess the skills required by these platforms or tools such as, a great deal of programming

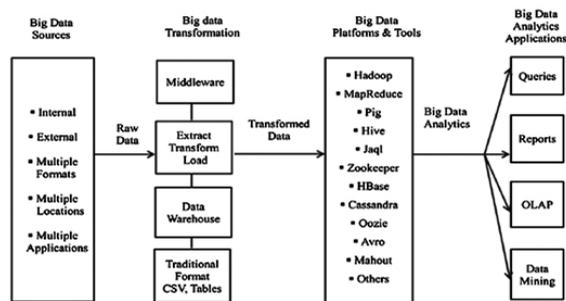


Figure 3: Architecture of Big Data Analytics

T. Murdoch, A. Detsky, (2013)

#### 6. Overall Goals of Big Data Analytics in Healthcare

- Take advantage of the massive amounts of data and provide right intervention to the right patient at the right time.
- Personalized care to the patient.
- Potentially benefit all the components of a healthcare system i.e., provider, payer, patient, and management.

#### 7. Need for Big Data Analytics in Healthcare

To improve the quality of healthcare by considering the following:

- Providing patient centric services.
- Detecting spreading diseases earlier.
- Monitoring the hospital's quality
- Improving the treatment methods.

#### 8. Big Data Challenges in Healthcare

There are many challenges that must be addressed to realize the full potential of big data. Provide a comprehensive discussion of such challenges based on the notion of data analysis: (H.V. Jagadish et al., 2012).

- Data Acquisition and Recording.
- Information Extraction and Cleaning.
- Data Integration, Aggregation and Representation.
- Query Processing, and Analysis.
- Efficiently handling large volumes of medical data and extracting potentially useful information and biomarkers.
- Analyzing genomic data is a computationally intensive task and combining with standard clinical data adds additional layers of complexity.

#### 9. Implementation of Big Data Analytics (BDA) Platform in Healthcare

Healthcare authorities and hospital systems need BDA platforms to manage and derive value from existing and future datasets and databases. Under the umbrella of the hospital system with its end users, the BDA platform should harness the technical power and advanced programming of accessible tools to analyze large quantities of data in an interactive manner, while enriching the user experience with data visualizations. All this must be accomplished at moderate expense for a successful platform to be used.

#### Methodology

While several different methodologies are being developed in this rapidly emerging discipline, here we outline one that is practical and hands-on.

In Step 1 develops a 'concept statement'. This is a first cut at

establishing the need for such a project. The concept statement is followed by a description of the project's significance.

Step 2, the proposal development stage. Here, more details are filled in. Based on the concept statement, several questions are addressed:

- What problem is being addressed?
- Why is it important and interesting to the healthcare provider?
- What is the case for a 'big data' analytics approach?

(Because the complexity and cost of big data analytics are significantly higher compared to traditional analytics approaches, it is important to justify their use).

Next, in Step 3, the concept statement is broken down into a series of propositions. The independent and dependent variables or indicators are identified. The data sources are also identified; the data is collected, described, and transformed in preparation for analytics. A very important step at this point is platform tool evaluation and selection. There are several options available, as indicated previously, including Hadoop, spark, and IBM Big Insights. The next step is to apply the various big data analytics techniques to the data. This process differs from routine analytics only in that the techniques are scaled up to large data sets. Insight is gained from the big data analytics. From the insight, informed decisions can be made.

In Step 4, the models and their findings are tested and validated and presented to stakeholders for action. Implementation is a staged approach with feedback loops built in at each stage to minimize risk of failure.

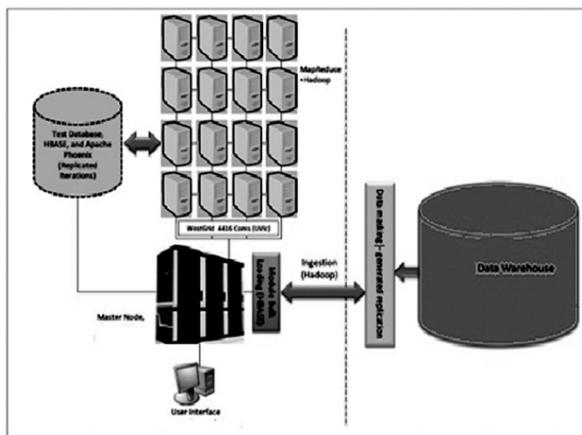


Figure 4: The proposed Healthcare Big Data Analytics (HBDA) framework

## 10. Apache Hadoop

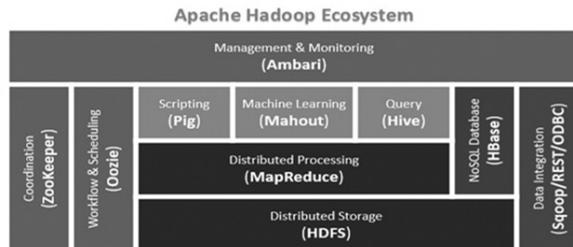


Figure 5: Apache Hadoop Ecosystem (P. Vignesh Raja and E. Sivasankar,2014)

- Apache Hadoop is one of the main supportive elements in Big Data technologies. It simplifies the processing of large amount of structured or unstructured data in a cheap manner.
- Hadoop is an open source project from apache that is continuously improving over the years.
- "Hadoop is basically a set of software libraries and frameworks to manage and process big amount of data from a single server to thousands of machines.
- It provides an efficient and powerful error detection mechanism based on application layer rather than relying upon hardware."

Apache Hadoop has two main subprojects:

### 1. Map Reduce

- It is basically a framework to write applications that processes a large amount of structured or unstructured data over the web.
- By distributed query processing it makes it easy to maintain large amount of data by dividing the data into several different machines.
- Hadoop MapReduce is a software framework for easily writing applications to manage large amount of data sets with a highly fault tolerant manner.

### 2. HDFS (Hadoop distributed file system)

- HDFS is a java based file system that is used to store structured or unstructured data over large clusters of distributed servers.
- The data stored in HDFS has no restriction or rule to be applied, the data can be either fully unstructured or purely structured.
- In HDFS the work to make data sensual is done by developer's code only.

## RESULTS

We have carried out experiments on healthcare data in Open source web interface HUE 3.7.1. It is a web interface for ana-

lyzing data with Apache Hadoop. The Healthcare dataset related to hospitals general information, inpatient, outpatient, type of diseases patients having etc. The dataset is demonstrated in pig editor using Hue interface. In these experiments, we have analyzed hospital general information such as general complications, diseases and hospital information. We try to experiment on diseases, which are affected by the patients and in what kind of hospital he is going to join. We also experimented on what types of complications in each hospital, across hospitals etc., have performed. The below Figure 6 represent the analysis of complications of hospitals using HUE.



Figure 6: Analysis of Hospitals dataset using Hue Pig Editor

### Conclusions

This paper discusses the role of Big Data in today's healthcare sectors. The transformation of using sophisticated technologies by healthcare providers to gain insights from clinical datasets and make informed decisions had changed by Big Data Analytics. With the help of Hadoop, the goal of effective healthcare management can be achieved by providing effective data driven services to people by predicting their needs. In future, Big Data is going to transform medical sectors across all healthcare organizations and healthcare industry by using Big Data Analytics. Rapid advances in applications of healthcare data, there is an acute need of Big Data Analytics. Required tools and practices can accelerate their progress in process and provides the results for analyzing massive datasets of healthcare domain along with required solutions. We hope this paper gives an admiring impact that would provide enough context and meaningful information about the approaches to adopt Big Data Analytics in healthcare sector.

### References

1. Adams J and Klein J., (2011). "Business Intelligence and Analytics in Health Care - A Primer", USA Springer.
2. Anonymous, (2012). "The value of analytics in healthcare - From insights to outcomes". Software Engineering (ICSE), International Conference; Vol. 9, No. 7.
3. Dembosky A, (2012). "Data Prescription for Better Healthcare." Financial Times.

4. Feldman B, Martin EM, Skotnes T, (2012). "Big Data in Healthcare Hype and Hope.", Journal of Medical Systems Vol.8 No. 2.
5. Ferrand, D., (2010). Towards a business intelligence framework for healthcare safety. Journal of Internet Banking and Commerce.
6. Frye, G.W., (2010). "Using business intelligence to build optimal decision support. Benefits & Compensation Digest".
7. H.V. Jagadish et al., (2012). "Challenges and Opportunities with BigData", available at <http://cra.org/ccc/docs/init/bigdata-whitepaper.pdf>
8. Issa T. and Isaias P., (2012). "Artificial Intelligence Technologies and the Evolution of Web 3.0", IGI Global.
9. Jinpon, P., Jaroensutansinee, M., & Jaroensutansinee, K., (2011). "Business intelligence and its applications in the public healthcare system". Walailak Journal of Science and Technology Vol.2 No. 03 Pages: 30-46.
10. Miller K, (2012). "Big Data Analytics in Biomedical Research", Biomedical Computation Review.
11. Mukherjee, A. Datta, J. Jorapur, R. Singhvi, R. Haloi, S. Akram, W. (2012). Shared disk big data analytics with Apache Hadoop, High Performance Computing (HiPC), 19th International Conference.
12. PwC Australia, (2012), Big Data - The next frontier for innovation, available: <http://www.pwc.com.au/consulting/assets/publications/Big>
13. Picciano, A. G., (2012). "The Evolution of Big Data and Learning Analytics in American Higher Education". Journal of Asynchronous Learning Networks Vol.2 No. 04
14. Ribeiro, R., (2006). Intelligent Decision Support Tool For Prioritizing equipment Repairs in Critical/Disaster Situations. Zaraté.
15. Russom, P. (2011). "Big Data Analytics (Best Practices)". The Data Warehouse Institute (TDWI).
16. Russom, P., (2011). TDWI Best Practices Report: Big Data Analytics (Best Practices). The Data Warehouse Institute (TDWI).
17. Siemens, G., & Long, P (2011). Penetrating the Fog: "Analytics in Learning and Education".
18. T. Murdoch, A. Detsky, (2013). "The Inevitable Application of Big Data to Health Care", JAMA, 2013, 309(13):1351-1352.
19. Wullianallur and Viju, (2014). "Big data analytics in healthcare: promise and potential", Health Information Science and Systems
20. Weiyi Shang, Zhen Ming Jiang, Hemmati, H. Adams, B. Hassan, A.E. Martin, P. (2013). "Assisting developers of Big Data Analytics Applications when deploying on Hadoop clouds", Software Engineering (ICSE), 35th International Conference.