

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 12, December 2019

## Designing Big Data Service for Ethiopian Healthcare System

Lijalem Megibaru Eneyew

School of Computing, Debre Markos Institute of Technology, Debre Markos, Ethiopia

**ABSTRACT:** Now a days the healthcare information needs to be organized for further processing and knowledge extraction. Regarding this implementing the big data analytical technologies will be a vital solution to collect, clean, model and deliver the results which will be used to make decisions and to plan for the future. All Ethiopian healthcare institutions will be given a privilege to push real time data to the big data cloud. So that scientists and researchers will play with the data that will help the government to make decisions and take actions based on the result. This paper highlights the enormous impact of big data in the healthcare industry particularly in Ethiopia.

**KEYWORDS:** Big Data, Big Data Analytics, Healthcare

### I. INTRODUCTION

In healthcare industries, large data is generated which is related to the patients record and other issues. So, in the age of modern technology, it is better to keep all the data in digitized form. Indirectly it has the positive impact on the improvement of the healthcare system at all. 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 [1]. This paper provides an overview of big data analytics in healthcare and government systems. It describes about big data generated by these systems, data characteristics, security issues in handling big data and how big data analytics helps to gain a meaningful insight on these data set. A Big Data analysis is a process that organizes the data collected from various sources and then analyzes the data sets in order to discover the facts and meaningful patterns [2].

Large-scale data analyzes have many uses in the field of health: for example, early diagnosis of diseases such as breast cancer, in the processing of medical images and medical signals for providing high-quality diagnosis, monitoring patient symptoms, tracking chronic diseases such as diabetes, preventing incidence of contagious diseases, education through social networks, genetic data analysis and personalized (precision) medicine. Therefore, analysis of health-related data has been considered with regard to its importance, so that it has led scholars and scientists to create structures, methodologies, approaches and new approaches for managing, controlling and processing Big Data [3].

### II. BIG DATA IN HEALTHCARE

In addition to the Volume, Velocity, Variety, Veracity, Vagueness, Versatility features, Big Data in health care has its own unique features, such as heterogeneity, incompleteness, timeliness and longevity, data privacy, and ownership. Defining the volume of the data, the type of data, and the entity limitations are very wide. Combining the complete healthcare data is very large, which refers as big data but not ashuge, as we think that the organization cannot handle its data. Most of the healthcare providers have not faced any hardersituation to handle the data for them; however, it is alwaysgood to predict the technology improvements andimplementation, which can help them. Hadoop dataprocessing is the one of the best choices to go with at thecurrent trends. The computational capabilities of Hadoopprocessing will be able to stimulate the mathematical methodsavailable currently, medical research

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 12, December 2019

approaches to increase the outcome quality. Although big data might not be their case, but the discovery process to find new techniques to analyze the data they have, increase the accuracy of the experimental results, provide various mechanisms to find the quality of data always at the highest priority. As a result, Hadoop data processing is one of the best choices to go with the current trends. The computational capabilities of Hadoop processing will be able to stimulate the mathematical methods available currently, and medical research approaches to increase the outcome quality. [3]

**Big Data as a Service (BDaaS):** In this digital world the amount of data processed increases rapidly. Use of social media and smart devices produce large amounts of data. This large amount of data needs management to control it. These large datasets are popularly known as Big Data. This big data is difficult to manage using traditional computing technologies [4]. New technologies are emerging in the market to address the problem of managing and analyzing Big Data. Organizations are getting it difficult to implement these Big Data technologies effectively due to problems lack of available expertise [5]. Big Data-as-a-Service provides common big data related services to users to enhance efficiency and reduce cost.

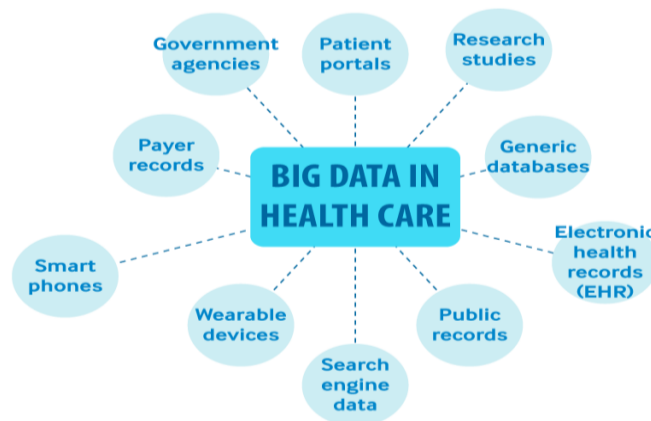


Fig 1. Sources of big data in healthcare [6]

Big Data Analytics in Healthcare has multiple significant. Some of the advantages are:

**Detecting spreading diseases earlier:** Predicting the viral diseases earlier before spreading based on the big data analysis. This can be identified by analyzing the social logs of the patients suffering from a disease in a particular geographical location. This helps the healthcare professionals to advise the victims by taking necessary preventive measures.

**Research purpose:** Researchers can access the data to analyse for further investigation. So that they will contribute the result for decisions to be made in that specific scenario.

**Providing patient centric services:** To provide faster relief to the patients by providing evidence-based medicine detecting diseases at the earlier stages based on the clinical data available, minimizing drug doses to avoid side effect and providing efficient medicine based on genetic makeup [4]. This helps in reducing readmission rates thereby reducing cost for the patients.

**Monitoring the hospital's quality:** Monitoring whether the hospitals are setup according to the standards.

**Improving the treatment methods:** Customized patient treatment monitoring the effect of medication continuously and based on the analysis dosages of medications can be changed for faster relief. Monitoring patient vital signs to provide proactive care to patients. Making an analysis on the data generated by the patients who already suffered from the same symptoms, helps doctor to provide effective medicines to new patients. [5]

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 12, December 2019

Predicating the future: if we use big data analytics algorithms to study the trend, we can predict the future and can suggest the measurements to be taken based on the past experience.

Healthcare is among the sectors generating a massive amount of data characterized by its high velocity and variety such as laboratory data, medical prescriptions, appointments, machine generated data, insurance data, and administrative data. [7]

Generally big data in healthcare will be used for Healthcare monitoring, Healthcare prediction, Performance enhancement, Healthcare knowledge system, Recommendation systems and Healthcare management system.

### III. LITERATURE REVIEW

The vast amount of data available in the health industry has become more influential in healthcare so that big data enables health systems to use this challenge as an opportunity to provide better analytical solutions. The increasing volume of healthcare data will be captured in a central repository which is accessible through the big data cloud. This data in the healthcare environment should be adopted by big data systems and technologies to effectively collect, analyze, and take advantage of the valuable information in the database. Many researches have been done so far related to the contribution of big data in the healthcare industry. The papers are trying to address the benefits that will be gained by deploying big data in the health care [5], the conceptual big data architectures that will be described based on different hierarchical layers [12], the tools and analytical methods most appropriately used to implement the big data system [6] and the challenges that must be considered while using big data [3].

This paper focuses mainly in providing big data service/platform to be deployed mainly in Ethiopian ministry of health which will be used for integrating all the health sector data in the country together with in a single Hadoop cloud further used to implement different analytical methods to extract valuable results from the analysis. This result further will be used to make decisions and to predict the country's health related plans.

So, I suggest the high-level big data architectural design that the government should apply to solve the problem of researchers and scientists to access the heterogeneous data within a single repository for their analytical facilities.

### IV. PROPOSED BIG DATA ARCHITECTURE

The ministry of health in Ethiopia is responsible to control the big data system centrally. The institution will have its own big data center. This Hadoop based cloud will be accessible for all granted researchers, scholars and students. The data will be pushed to the data center from all healthcare environments like Hospitals, health centers, clinics and other sectors which are giving healthcare services. This data from each of the healthcare industries will be further processed in the data center. In this big data as a service technology we are deploying effective analytical algorithms so that the result helps the organization to make decisions, to take actions, to analyze the trend or to predict based on the past experiences in the big data. The high-level architecture to be used in the analytics process have different layers.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 12, December 2019

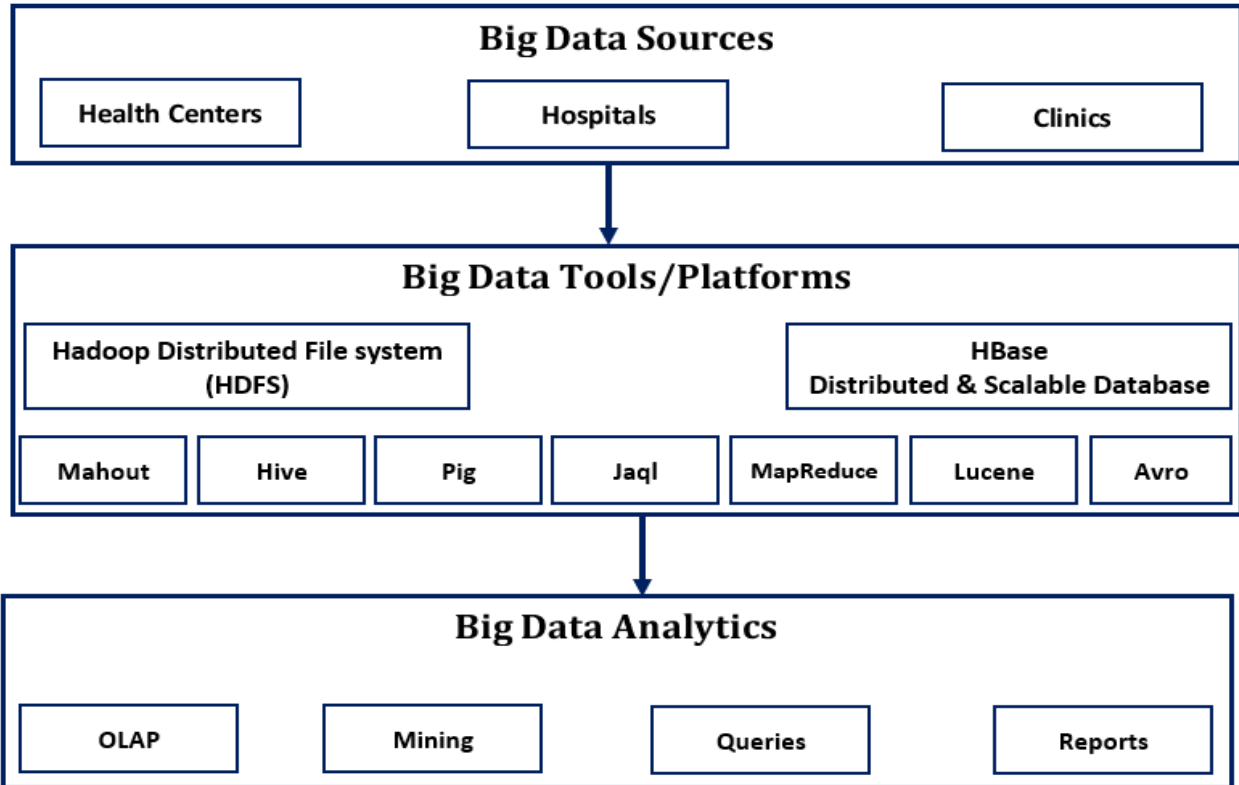


Fig 2. Big data analytics architecture

In this research there are platforms and tools used for big data analytics in the healthcare system:

**The Hadoop Distributed File System (HDFS):** HDFS enables the underlying storage for the Hadoop cluster. It divides the data into smaller parts and distributes it across the various servers/nodes.

**MapReduce:** MapReduce provides the interface for the distribution of sub-tasks and the gathering of outputs. When tasks are executed, MapReduce tracks the processing of each server/node.

**PIG and PIG Latin (Pig and PigLatin):** Pig programming language is configured to assimilate all types of data. It is comprised of two key modules: the language itself, called PigLatin, and the runtime version in which the PigLatin code is executed.

**Hive:** Hive is a runtime Hadoop support architecture that leverages Structure Query Language (SQL) with the Hadoop platform. It permits SQL programmers to develop Hive Query Language (HQL) statements akin to typical SQL statements.

**Jaql:** Jaql is a functional, declarative query language designed to process large data sets. To facilitate parallel processing, Jaql converts high-level queries into low-level queries consisting of MapReduce tasks.

**Zookeeper:** Zookeeper allows a centralized infrastructure with various services, providing synchronization across a cluster of servers. Big data analytics applications utilize these services to coordinate parallel processing across big clusters.

**HBase:** HBase is a column-oriented database management system that sits on top of HDFS. It uses a non-SQL approach.

**Lucene:** The Lucene project is used widely for text analytics/searches and has been incorporated into several open source projects. Its scope includes full text indexing and library search for use within a Java application.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 12, December 2019

Avro: Avro facilitates data serialization services. Versioning and version control are additional useful features.

Mahout: Mahout is yet another Apache project whose goal is to generate free applications of distributed and scalable machine learning algorithms that support big data analytics on the Hadoop platform. [8]

The analysis process in the big data will pass through different lifecycles in the proposed big data architecture. So, starting from the collection of the actual data to be stored/used in the big data until it produces the expected outcome, it has multiple processes to be done.

**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.

**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 Modelling:** It involves performing analysis on the classified data. For example, Government may require the list of malnourished children in a particular location. First it has to classify the data based on the specific location, need to trigger the health report of children, need to identify the children whose family are under poverty line and these data should be processed.

**Data Delivery:** It involves the generation of report based on the data modelling done. Based on the example after the data is processed it will generate a report based on malnourished children in a particular location. This will help the government to take necessary measures to avoid any further complications [9].

In all the stages of Big Data Lifecycle it requires data storage, data integrity and data access control.

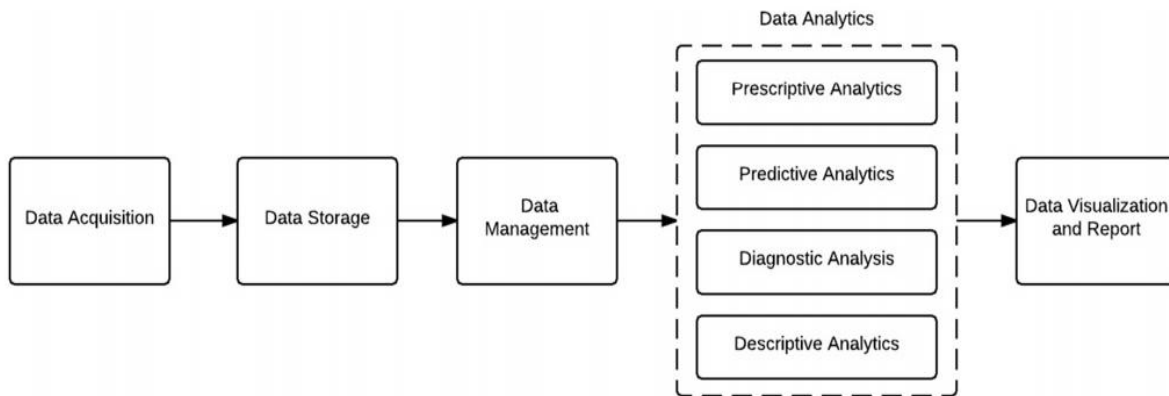


Fig 3. Big Data Lifecycle [3]

## V. CHALLENGES OF BIG DATA IN HEALTHCARE SYSTEMS

Healthcare centered big data has many challenges in the process of capturing, storing, sharing, searching, and analyzing health related data. The organization of data after extraction from different layers and the integration of the data, errors from analytical practice and experiment and Integration of physiological data with high-throughput techniques for clinical recommendations are all challenges of the big data process. The issue on agreement to using healthcare data such as genetic data has been a concern. Creating databases based on large and national population for future research with ethics approval and governance has led to academic arguments on legality. There are even arguments on that Big Data is useful to improve healthcare systems [10]. The following are general challenges of Big Data in healthcare [11]:

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 12, December 2019

- **Security and privacy:** Traditional privacy and security measures work on small datasets; capability to use the same measures on massive and streaming datasets is possibly a problem, particularly when dealing with patient's health data.
- **Data quality:** It affects reliable insights from the data and decision-making for patients' healthcare.
- **Insufficient real-time processing:** Delay in processing complex data models can result in patient care with less quality.
- **Integration of heterogeneous data sources:** data fragmentation across hospitals, labs, electronic health records (EHRs), and financial IT systems is a major obstacle to combining data into an integrated database system.
- **No fixed standards for healthcare data:** Many kinds of healthcare data are generated and collected by various agents such as practitioners' notes, medical images, data from wearable sensors. There are no unified standards for these data, which brings difficulty for further processing. [12]

## VI. CONCLUSION

In Ethiopia the problem is not the shortage of health-related data but the deficiency of information that can be used to support decision making and prediction-based planning in the health industries. The sector must consider the aids from applying big data technologies in the health industry to enhance the decision making, planning and strategies. To effectively identify and implement big data solutions and benefit from the value that big data can bring, sector need to allocate enough budget, facilities and resources for the big data technology particularly in the healthcare industry. With the support of big data technology, specially Hadoop, the goal of effective health care management can be achieved by providing an actual data driven services. Secured Big data analytics must be employed in a security enabled platform.

Generally, it is highly recommended that the Ministry of Health better to introduce the big data technology to create central Hadoop cloud system to manage the data in effective and organized way for decision making.

## REFERENCES

1. W. Raghupathi, V. Raghupathi Big data analytics in healthcare: promise and potential Health Inf Sci Syst, 2 (1) (2014)
2. M. Kasem, E. Hassanein Cloud business intelligence survey Int J Comput Appl, 9 (1) (2014)
3. R. Fang, S. Pouyanfar, Y. Yang, C. Chen Computational health informatics in the big data age: a survey ACM Comput Surv, 49 (1) (2016), pp. 1-36  
Elsevier, USA (2011)
4. K. Kalpana, L. Bansal Comparative study of data mining tools Int J Adv Res Comput Sci Softw Eng, 4 (6) (2014)
5. A. Linden, P. Krensky, J. Hare, C. Idoine, S. Sicular, S. Vashisth Magic quadrant for data science platforms Gartner (2017)
6. E. Siegel Predictive analytics: the power to predict who will click, buy, lie, or die  
Wiley & Sons, New Jersey (2013)
7. C. Duhigg How companies learn your secrets The New York Times Magazine (2012)
8. R. Whiting Businesses mine data to predict what happens next InformationWeek (2006)
9. Knoppers, B. M., & Thorogood, A. M. (2017). Ethics and big data in health. *Current Opinion in Systems Biology*, 4, 53-57.
10. Mathew, P. S., & Pillai, A. S. (2015, March). Big data solutions in healthcare: problems and perspectives. In *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on* (pp. 1-6). IEEE.
11. A. McAfee, E. Brynjolfsson Big data: the management revolution Harv Bus Rev, 90 (10) (2012),
12. J. Admes, D. Garets the healthcare analytics evolution: moving from descriptive to predictive to prescriptive R. Gensinger (Ed.), *Analytics in healthcare: an introduction*, Health Information and Management System Society (HIMSS), Chicago (2014), pp. 13-20