

# Big Data and Data-Driven Healthcare Systems

Cheryl Ann Alexander<sup>1,\*</sup>, Lidong Wang<sup>2</sup>

<sup>1</sup>Institute for Information Technology Innovations & Smart Healthcare, Vicksburg, MS, USA

<sup>2</sup>Institute for Systems Engineering Research, Mississippi State University, Vicksburg, MS, USA

\*Corresponding author: [tech.health.sol@gmail.com](mailto:tech.health.sol@gmail.com)

**Abstract** Data analytics has been used in healthcare. Healthcare systems generate big data. Traditional data management techniques are often unable to manage the voluminous amounts of data produced in healthcare systems. Big Data analytics which is overcoming the limitations of traditional data analytics will bring revolutions in healthcare systems. Big data and Big Data analytics in healthcare systems are presented in this paper. Information security, privacy, and challenges of Big Data analytics in healthcare are also discussed.

**Keywords:** *big data, big data analytics, cloud computing, fog computing, stream processing, healthcare, personalized medicine, information security, privacy*

**Cite This Article:** Cheryl Ann Alexander, and Lidong Wang, "Big Data and Data-Driven Healthcare Systems." *Journal of Business and Management Sciences*, vol. 6, no. 3 (2018): 104-111. doi: 10.12691/jbms-6-3-7.

## 1. Introduction

Medical science has long relied on clinical trials to demonstrate the efficacy of interventions, whether pharmaceutical, surgical, or device based. Medical interventions should frequently be tailored to the specific characteristics of each individual patient. There has been an increased focus on personalized medicine in recent years, which relies on tailoring to individuals to provide "the right drug for the right patient at the right dose and time" [1]. Electronic health records and the data automatically collected from devices such as wearable devices are often sources of big data. It is not easy to perform perfect data curation and quality control. While approaches driven by Big Data accelerate the discovery of new therapies and diagnostics, all computational predictions must still be thoroughly validated in experimental and clinical settings before widespread use. People are moving toward big data-based healthcare, including data-driven methodologies to accelerate the discovery of new diagnostics and drugs [2].

Some data related to healthcare is characterized by a need for timeliness, such as data from implantable or wearable biometric sensors or the heart rate or SpO<sub>2</sub> which is commonly gathered and analyzed in real time. Suitable large-scale analysis typically requires the gathering of data from numerous sources (or heterogeneous data), for example, obtaining a patient's (or a population's) comprehensive health status requires the integration and analysis of patient health records beyond Internet-available, environmental data, or assorted meter readings (e.g., accelerometers, remote, wearable, or local cardiac monitors, or glucometers) [3]. Big Data approaches are being used to build models of healthy aging. Age-related conditions are the leading causes of death and healthcare costs. Reducing the rate of aging would have enormous

medical and financial benefits. Myriad genes and pathways are known to regulate aging in model organisms. Challenges and pitfalls of commercialization include reliance on findings from short-lived model organisms, poor biological understanding of aging, and hurdles in performing clinical trials for aging [4].

Analytics in healthcare is driven by the gradual shift from disease-centered to patient-centered care (PCC). From a general practitioner's desktop computer to cardiac monitors in an emergency room, a multitude of clinical information systems capture patient information. This information exists at different levels of granularity, in diverse formats and recorded at varying frequency. A patient can record blood glucose levels at different times during the day when at home whereas a clinic may capture a single measurement but derive a different measure (glycated haemoglobin) to determine the three months average. This difference in granularity can be an extra dimension of information for Big Data analytics (BDA) when paired with medication, demographic or behavioral information. BDA can be used to identify changes in medical images and relate these to changes in medications or dosage. The inherent limitations of most data collections, such as missing data, null values, incorrect values and unmatched records were observed and accounted for in the BDA process. The fusion of structured and unstructured data is aptly demonstrated in the outcomes and is of significant value in a clinical context. BDA architecture for healthcare applications should overcome the complexities of granular data accumulation, temporal abstraction, multimodality, unstructured data and integration of multi-source data to provide a robust platform for effective workflows and improved engagement [5].

The variety of big data is not solved only by parallelizing and distributing problems. Variety is mitigated by capturing, structuring, and understanding unstructured data using artificial intelligence (AI) and other different analytics [6]. Clinical data is expressed within the

narrative portion of the EMRs, requiring natural language processing techniques to unlock the medical knowledge referred to by physicians [7]. Research on big data has mostly focused on addressing technical issues. However, organizations will not acquire the full benefits of leveraging big data analytics unless they can address managerial challenges effectively, orchestrate strategic choices and resource configurations, as well as understand the managerial, economic, and strategic impact of big data analytics. Moving a deeper understanding on the ways and means to create business value from big data analytics will result in reducing a resistance to adopt big data analytics and an ineffective use of analytics. Thus, exploring the path to big data analytics success for healthcare transformation is currently one of the most discussed topics in the fields of computer science, information systems (IS), and healthcare informatics [8].

## 2. Data Sources in Healthcare and Big Data Advantages

Healthcare datasets collected in both clinical and non-clinical segments are in various forms and their source are described in Figure 1. Some keywords related to big data in the biomedical area are listed in Table 1. Big-Omic Data are the data containing a comprehensive catalog of molecular profiles (e.g., genomic, transcriptomic, epigenomic, proteomic, and metabolomic in biological samples that provide a basis for precision medicine. Big EHR Data can be unstructured (e.g., clinical notes) or structured (e.g., ICD-9 diagnosis codes, administrative data, chart, and medication). Omic and EHR big data analytics is a challenge due to data frequency, quality, dimensionality, and heterogeneity [9].

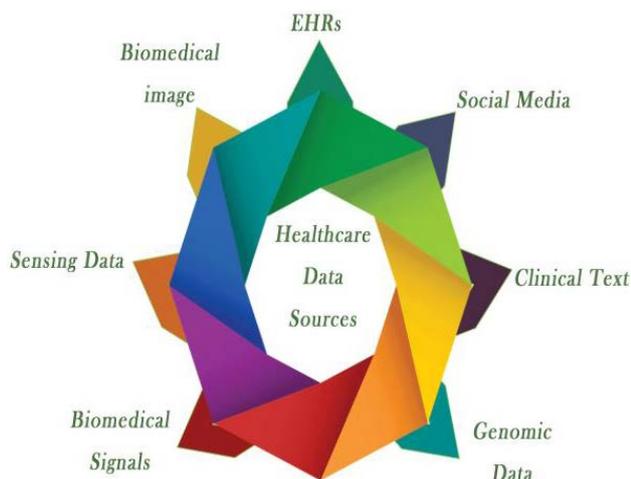


Figure 1. Data sources of the healthcare system [10]

Developing a detailed model of a human being by combining physiological data and high-throughput -omics techniques has the potential to enhance the knowledge of disease states and help develop blood-based diagnostic tools. Medical image analysis, signal processing of physiological data, and integration of physiological and -omics data face challenges and opportunities in dealing with disparate structured and unstructured big data sources [11]. Big data technologies are increasingly used for processing next-generation sequencing (NGS) data, motivated by the volume and velocity at which sequencing data is produced. Existing implementations of cloud-enabled NGS tools often use the MapReduce (MR) paradigm. MR is included in frameworks such as Hadoop that enable distributed processing of large-scale NGS datasets on a cloud [12].

Infectious disease surveillance is one of the most exciting opportunities created by big data because these novel data streams can improve timeliness, spatial and temporal resolution. These streams can also go beyond disease surveillance and provide information on behaviors and outcomes related to vaccine or drug use [13]. Big Data can be used in the health care to get innovative outcomes in the following areas [14]:

- Public and population health: BDA solutions can mine web-based data and social media data to predict the trend of diseases (e.g. flu).
- Evidence-based medicine: it involves the use of statistical studies and quantified research by doctors to form diagnosis.
- Clinical decision support: BDA technologies can be used to predict outcomes or recommend alternative treatments to clinicians and patients at the point of care.
- Personalized care: predictive data mining or analytic solutions may offer early detection and diagnosis before a patient has disease symptoms. Pattern detection can be fulfilled through real time wearable sensors for elderly or disabled patients to alert the physicians if there is any change in their vital parameters or post-market monitoring of drug effectiveness.
- Fraud Detection: fraud in medical claims can increase the burden on the society, Predictive models like decision tree, neural networks, regression etc. can be used to predict and prevent fraud at the point of transactions.
- Secondary usage of health data: dealing with aggregation of clinical data from finance, patient care, administrative records to discover valuable insights like identification of patients with rare disease, therapy choices, clinical performance measurement etc.

Table 1. Keywords related to big data in the biomedical area [9]

Topics	Keywords
EHR Data	Clinical data management, medical coding systems, next-generation EHR, Big data in HER, etc.
EHR Data Analytics	Clinical decision support, Temporal medical data mining, irregular time series analysis in EHR, clinical, waveform analysis in EHR, supervised/ unsupervised learning in EHR, etc.
-Omic Data	Genomics, proteomics, metabolomics, transcriptomics, epigenomics, etc.
-Omic Data Analytics	-Omic dynamic modeling, -omic network analysis, MS preprocessing, NGS variant detection, NGS sequence mapping, NGS differential analysis, NGS biomarker identification, RNA-seq computation, ChIP-seq computation, etc.
Big Data Analytics Enablers	Big data framework, big data platform, big data harmonization,
Data Challenges	EHR data challenges, -omic data challenges, biomedical big data challenges, etc.

### 3. Case Studies of Big Data in Diseases

In diabetes, a multidimensional approach to data analysis is needed to better understand the disease conditions, trajectories and the associated comorbidities. Elucidation of multidimensionality comes from the analysis of factors such as disease phenotypes, marker types, and biological motifs while seeking to make use of multiple levels of information including genetics, omics, clinical data, and environmental and lifestyle factors. A significant role is played by both environmental and genetic factors in Type-2 diabetes (T2D) [15]. Predictive analysis algorithm was used in Hadoop/MapReduce environment to predict the diabetes types prevalent, complications associated with it and the type of treatment to be provided. The healthcare industry is moving from reporting facts to discovery of insights, toward becoming data-driven healthcare organizations. Big Data holds great potential to change the whole healthcare value chain from drug analysis to patients caring quality [16].

Medical images help in early detection, diagnosis and prognosis of neurological disorders. Diagnosis of this disease by the radiologists is achieved through the neuroimaging techniques. The major constituents of human brain are Gray Matter (GM), White Matter (WM)

and Cerebrospinal Fluid (CSF). Cranial volume is a significant metric by which the abnormality in the size and shape of brain is detected. Hence quantitative analysis of brain tissues plays a key role in the diagnosis of these illnesses. Performing this measurable analysis on MRI brain images from a medical imaging perspective requires image processing and/or machine learning techniques. On the other hand, understanding why there is loss of neurons is viewed from a bioinformatics perspective. Studies have confirmed that one of the reasons owes to protein misfolding where the proteins fail to fold appropriately. This leads to severe concerns resulting in neuronal death [17]. Big Data has a great potential in the study of brain science.

Figure 2 shows Big Data–driven discovery in gastroenterology and hepatology: 1) Big Data–driven discovery can provide new approaches to long-standing or emerging unmet needs in gastrointestinal and liver diseases; 2) systematically and/or automatically collected heterogeneous data from patients and publicly or privately available databases are integrated into a highly rich datasets and analyzed; 3) mining assembled big data by specialized methodologies (translational bioinformatics) efficiently yields diagnostic devices, tools, and/or therapeutics [2].

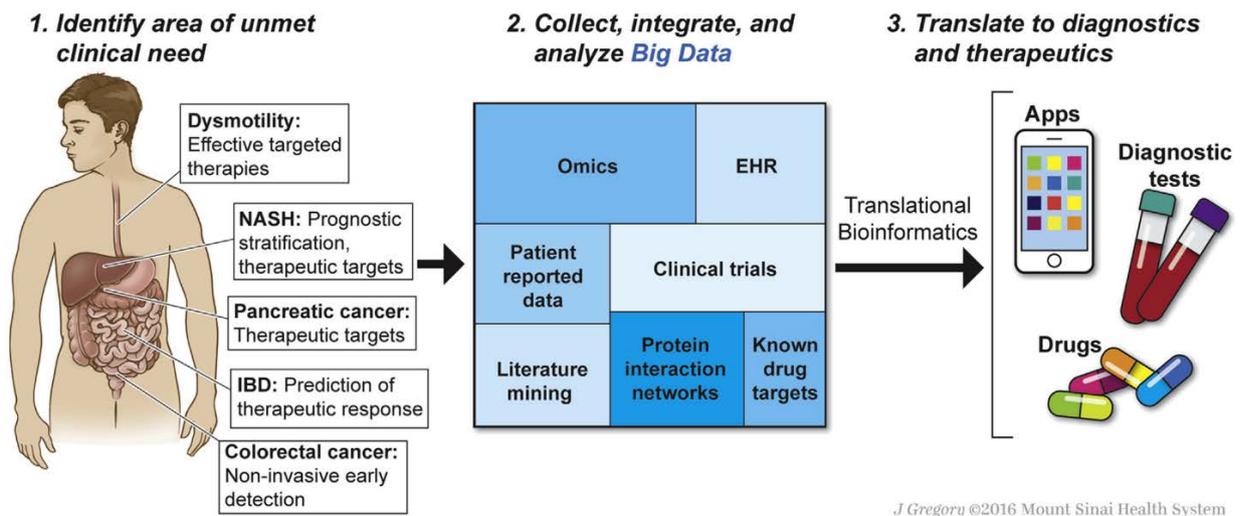


Figure 2. Big data–driven discovery in gastroenterology and hepatology [2]



Figure 3. Online Healthcare Monitoring System for Improving Analytics [19]

### 4. Advances of Big Data in Healthcare

Big data analytics comprises an integrated array of aggregation techniques, analytics techniques, and interpretation techniques that allow users to transform data into evidence-based decisions and informed actions. Data aggregation aims to collect heterogeneous data from multiple sources and transforming various sources data into certain data formats that can be read and analyzed. Data will be aggregated by three key functionalities from data aggregation tools: acquisition, transformation, and storage. Data analysis aims to process all kinds of data and perform appropriate analyses for harvesting insights. Data interpretation generates outputs such as various visualization reports, real-time information monitoring, and meaningful business insights derived from the analytics components

to users [18]. An online healthcare monitoring system was developed that is shown in Figure 3.

Figure 4 shows an advanced process of data collection. Various healthcare data are collected by data nodes and are transmitted to the cloud through configurable adapters that provide the functionality to preprocess and encrypt the data [20]. Figure 5 [22] shows the working flow of a healthcare monitor system based on the healthcare cloud, in which the webpage interface provides four basic query options, including real-time dynamics, status overview, device distribution, and patient-healthcare record. Fog computing (shown in Figure 6) is an emerging paradigm that provides storage, processing, and communication services closer to the end user. Fog computing does not replace cloud computing. Rather, it extends the cloud to the edge of the network [21].

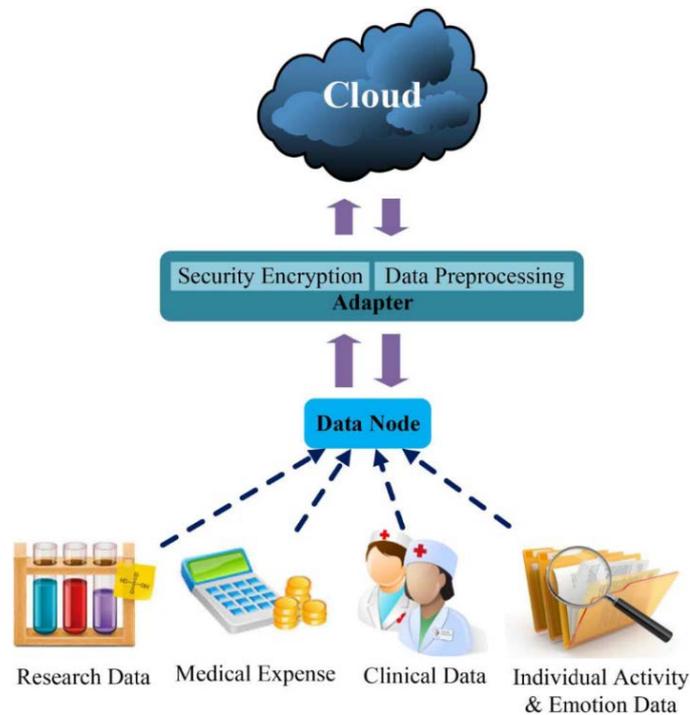


Figure 4. Data collection [20]

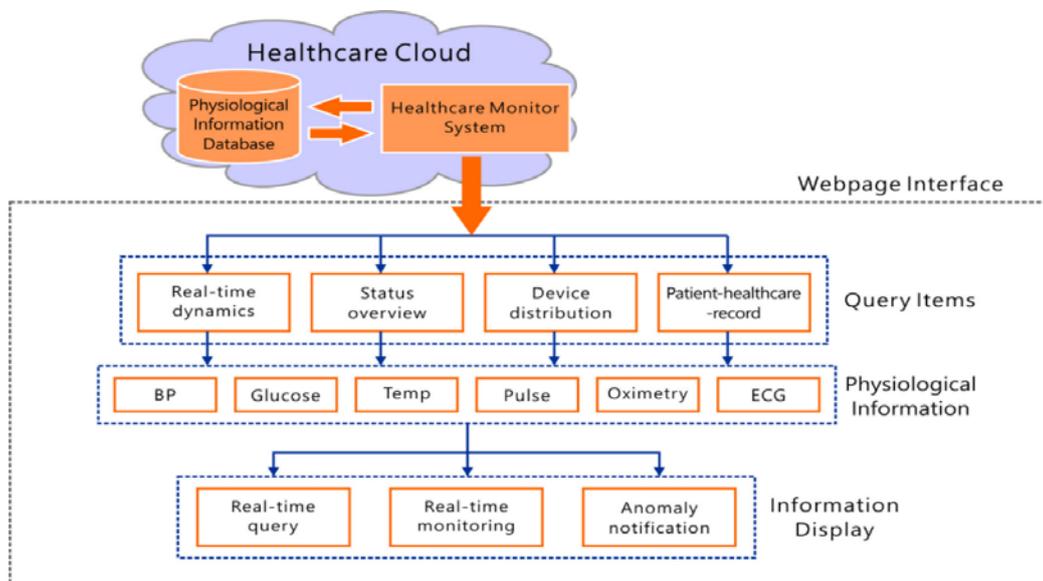


Figure 5. The working flow of a monitor system in healthcare

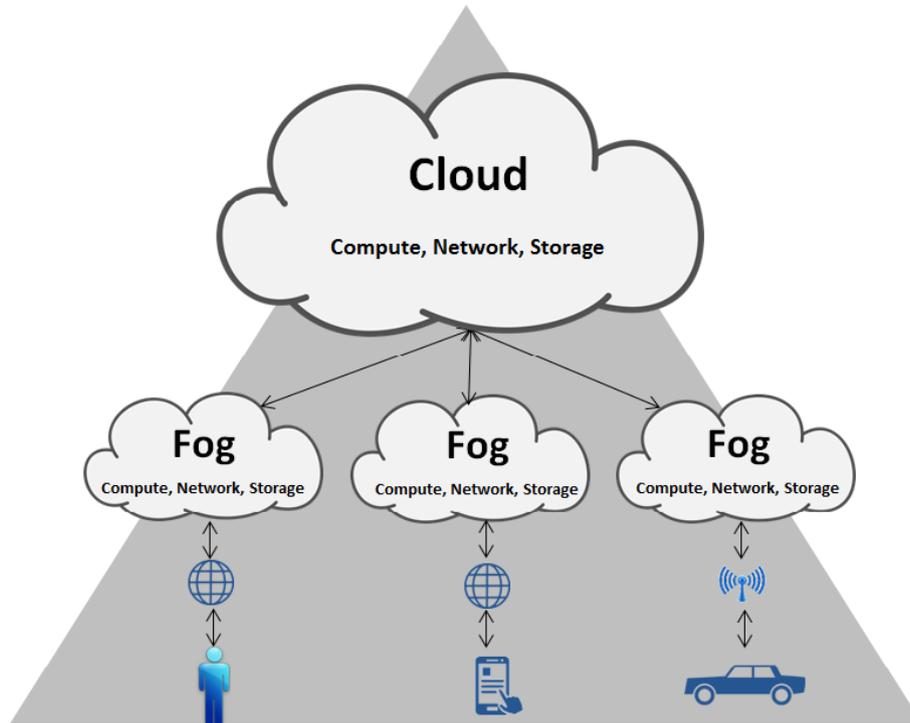


Figure 6. Fog computing architecture [21]

Big data computing is a new trend for future computing with a large amount of data sets and can be divided into three paradigms: batch-oriented computing, real-time oriented computing (or stream computing), and hybrid computing. Apache Hadoop (Hadoop) is an example of batch-oriented computing. However, the output time will vary depending upon the amount of data that is given as input. In contrast, real-time oriented computing involves continuous input and outcome of data. A big data input stream has three main characteristics namely high speed, real time, and large volume [10]. New technologies, such as platforms and infrastructures, are required for handling big data. A historical perspective of the frameworks of these technologies in data processing is shown in Figure 7 [23].

Researchers have presented a novel cloud platform for fast statistics and analysis based on big data processing technology. In this platform, medical service information is transformed to a new data structure in a column-oriented Data Base Manage System (DBMS); Spark cluster is used to satisfy the real-time computing requirements. Hadoop is one of the most important open-source big data platforms, and it simplifies the processing and management of big data by means of the MapReduce model and sophisticated ecosystem. The fast-statistical analysis platform composes of the following basic components: Data ETL Servers, Distributed storage, Spark cluster and Application Web Server. The fast computing becomes the critical step in the statistic and analysis of big medical service data. Therefore, it is imperative to use new big data processing platform for accelerating the computing and utilization of those medical service data. Fast statistical and analysis platform of medical service big data should provide the following functions: the design of new data structure in the distributed database(HBase) which is capable for processing hundreds of millions of records, transforming source-data from Oracle database to HBase database, executing statistics and analysis by using math methods and friendly output of computing results [24]. Table 2 describes the main components of the Spark framework.

Big data analytics includes the various analytical techniques such as descriptive analytics and mining/predictive analytics that are perfect for analyzing a sizeable quantity of text-based health documents and other unstructured clinical data (e.g., physician's written notes and prescriptions and medical imaging). Novel database management systems such as MongoDB, MarkLogic and Apache Cassandra for data integration and retrieval, allow data being transferred between traditional and new operating systems. To store the enormous volume and numerous formats of data, there are Apache HBase and

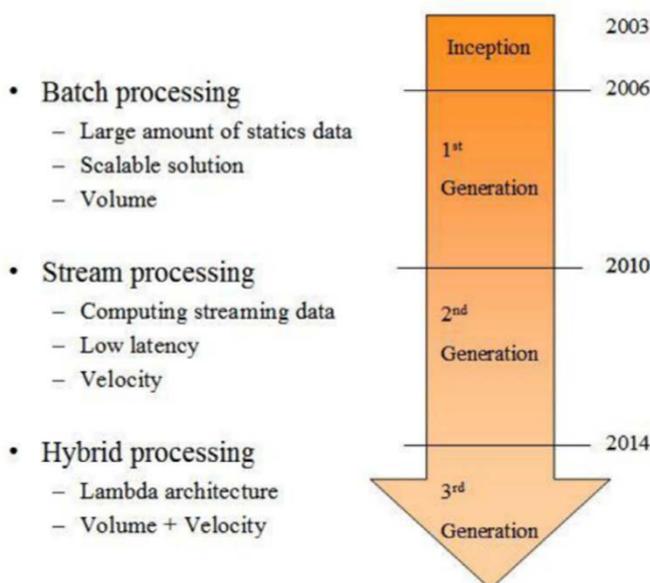


Figure 7. Three generations of data processing paradigms

NoSQL systems, which are tools with sophisticated functionalities that facilitate clinical information integration and provide innovative business visions [26]. New features of big data processing, such as insufficient samples, uncertain data relationships and unbalance (or even uncertain) distributions of value density, should be fully considered. Scalability and timeliness are two issues with high priorities regarding big data. The challenges of big data visualization come from the large sizes and high dimensions of data. Current visualization techniques mostly suffer from poor performances in functionalities, scalability and response time. Moreover, the effectiveness of visualization may be challenged by uncertainties of data sources [23]. As a successor of ipython, Jupyter was a successful interactive and development tool for data science and scientific computing. The HBDA platform

was developed and showed high performance tested for healthcare applications. With moderate resources, users are able to run realistic SQL queries on one billion records and perform interactive analytics and data visualization using Drill, Spark with Zeppelin or Jupyter. The performance times proved to improve over time with repeated sessions of the same query via the Zeppelin and Jupyter interfaces. An ingesting and using CSV file on Hadoop also had its advantages but was expensive when running Spark. Drill offers better low latency SQL engine but its application tool and visualization were very limited to customization; therefore, had lower usability for healthcare purposes [27]. A medical prototype was implemented in Centos64 operating systems. The distributed storage and Spark cluster are composed by 4 virtual machine nodes. The specific software configuration is shown in Table 3.

**Table 2. Spark Framework [25].**

Components	Description
MLib Machine Learning	Common library containing machine learning algorithms including clustering, classification, regression, etc.
GraphX Graph Processing	Framework for manipulating graphs
SparkSQL Structured Data	Package for handling structured data. Querying via SQL as well as Apache Hive. SparkSQL supports interface with RDBMS, HIVE Tables, NoSQL Databases, Parquet and JSON
Spark Streaming Real-Time	Supports streaming of live data. Spark makes use of Resilient Distributed Data (RDD) creation, processing and management subsystem
YARN MESOS Standalone Scheduler	Cluster Management Services
Spark Core	Contains basic Spark functionality. Sparks fundamental programming abstraction, RDD Set represents a collection of items spread across parallel computing nodes. Spark provided an API for creating and managing RDDs. This API also takes care of parallel processing and management of RDDs.

**Table 3. Software configuration of a medical prototype [24]**

Tools	Function	Description
Sqoop-1.4.3	Medical service data ETL	Transform data from Oracle to HBase
Hadoop2.4.0 HBase-0.98 Zookeeper-3.4.6	Distributed storage and cooperation	Medical service data and computation results storage in new data structures
Spark-1.2	Fast computation	Computation infrastructure
Web2py-1.95	Visualization	Web application

## 5. Security, Privacy and Challenges of Big Data in Healthcare

Medical data is highly sensitive and the federal Health Insurance Portability and Accountability Act (HIPAA) requires protecting the confidentiality and security of healthcare data. Various approaches have been developed based on privacy preserving data mining (PPDM) for protecting the privacy of individuals or groups within a dataset while maintaining the integrity of the knowledge contained within the data for knowledge discovery purposes. Sensitive data spanning multiple organizations result in not only data syntax and semantic heterogeneity but also diverse privacy requirements, which results in additional challenges to data sharing and integration. Data sharing for purposes such as billing and joint ventures is permissible under HIPAA regulations. Healthcare data

such as an electronic medical record (EMR) are valuable. Besides implications for patient privacy, a security breach has repercussions for healthcare providers such as diminished reputation, litigation, or imposed penalties [28].

The rise of big data generated by mobile sources has brought unprecedented opportunities for researchers to explore new possibilities. Opportunities presented by mobile big data (MBD) have been introduced. Mobility can amplify the effects of big data on both operational efficiency and customer intelligence by making everything instantly actionable, which can change business processes. However, solving MBD problems while respecting the privacy of customers is one of the biggest concerns of enterprises. MBD comprises personal location-based data, which users do not wish to reveal. Therefore, some research is required to identify new methods and technologies that can allow customers to dynamically

verify their data privacy according to the rules and regulations of their service level agreements. The development of such methods can ensure customer privacy. Without the proper assurance of privacy, enterprises may not be able to obtain complete data from customers, thereby possibly being misled in their decision making [29].

There are following challenges in the whole big data process [30]:

- Scalability. When we consider the integration of streams coming from all healthcare sport services with other IoT applications such as GPS sensors inside cars or air pollution sensors, the data flow can easily reach up to millions of tuples per second. Centralized servers cannot process flows of this magnitude in real time. Thus, the main challenge is to build a distributed system where every node has a local view of the data flow. These local views must then be aggregated to build a global view of the data with an off-line analysis.
- Heterogeneity and incompleteness. The IoT ecosystem generates heterogeneous data flows coming from different types of applications and devices. Therefore, the main challenge here is to integrate and structure massive and heterogeneous data flows coming from the IoT to prepare their analysis in real time.
- Timeliness. Speed in big data is important in both input and output. The input is represented by a huge dataset coming from multiples sources that must be processed and structured for analysis. The output is represented by results of analysis or queries over the dataset. The main challenge here is how to implement a distributed architecture that is able to aggregate local views of data inside every node into a single global view of results with minimal communication latency between nodes.
- Privacy. People generate and share personal data that are not always protected. Data generated from healthcare sport services contains sensitive personal information. A key challenge here is to propose techniques that protect this kind of data before its analysis.

The application of Software as a Service (SaaS) in healthcare domain is clearly a possible solution to handle large set of data on cloud. The available security measures help handle the data on cloud in a secured manner. Having a great service on the cloud that helps users to analyze the data from a remote location will be helpful for both patients and healthcare industry. This can solve overhead of people traveling to hospitals for every medical checkup [31]. Health Information Exchanges (HIEs) which support electronic sharing of data and information between health care organizations are recognized as a source of big data in healthcare and have the potential to provide public health with a single stream of data collated across disparate systems and sources. However, given these data are not collected specifically to meet public health objectives, it is unknown whether a public health agency's (PHA's) secondary use of the data is supportive of or presents additional barriers to meeting disease reporting and surveillance needs. The following challenges have been uncovered for effective utilization of big data by public health [32]:

- While PHAs almost exclusively rely on secondary use data for surveillance, big data that has been collected for clinical purposes omits data fields of high value for public health.
- Big data is not always smart data, especially when the context within which the data is collected is absent.
- Data collected by disparate, varying systems and sources can introduce uncertainties and limit trustworthiness in the data which may diminish its value for public health purposes.
- The process by which data is obtained needs to be evident in order for big data to be useful to public health.
- Big data for public health purposes needs to answer both 'what' and 'why' questions.

## 6. Conclusion

Data-driven management in healthcare systems has become a strategic choice in achieving sustainable growth, meeting the challenges of global competition, and explore the potential innovation for the future. Novel data analytics such as Big Data analytics are key to advancing healthcare systems. Big Data can be used in the health care to get innovative outcomes in public and population health, evidence-based medicine, clinical decision support, personalized care, fraud detection, etc. Artificial intelligence (AI) and Big Data analytics could reshape the healthcare systems with greater performance in productivity, efficiency, and the quality of care. Challenges in the big data process lie in scalability, heterogeneity and incompleteness, timeliness, privacy, etc.

## References

- [1] Price, W. N. (2015). Big Data, Patents, and the Future of Medicine. *Cardozo L. Rev.*, 37, 1401.
- [2] Wooden, B., Goossens, N., Hoshida, Y., & Friedman, S. L. (2017). Using big data to discover diagnostics and therapeutics for gastrointestinal and liver diseases. *Gastroenterology*, 152(1), 53-67.
- [3] Salman, O. H., Zaidan, A. A., Zaidan, B. B., Naserkalid, & Hashim, M. (2017). Novel methodology for triage and prioritizing using "big data" patients with chronic heart diseases through telemedicine environmental. *International Journal of Information Technology & Decision Making*, 16(05), 1211-1245.
- [4] Magalhães, J. P Stevens, M. and Thornton, D. (2017). The Business of Anti-Aging Science. *Trends in Biotechnology*, 35(11), 1062-1073.
- [5] De Silva, D., Burstein, F., Jelinek, H. F., & Stranieri, A. (2015). Addressing the complexities of big data analytics in healthcare: the diabetes screening case. *Australasian Journal of Information Systems*, 19.
- [6] Kalantari, A., Kamsin, A., Shamshirband, S., Gani, A., Alinejad-Rokny, H., & Chronopoulos, A. T. (2017). Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. *Neurocomputing*.
- [7] Harabagiu, S. (2015, December). Big mechanisms for processing big data in medical informatics. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2015 IEEE* (pp. 1-2). IEEE.
- [8] Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, 70, 287-299.
- [9] Wu, P. Y., Cheng, C. W., Kaddi, C. D., Venugopalan, J., Hoffman, R., & Wang, M. D. (2017). -Omic and Electronic Health Record

- Big Data Analytics for Precision Medicine. *IEEE Transactions on Biomedical Engineering*, 64(2), 263-273.
- [10] Vanathi, R., & Khadir, A. S. A. (2017, February). A Robust Architectural Framework for Big Data Stream Computing in Personal Healthcare Real Time Analytics. In *Computing and Communication Technologies (WCCCT), 2017 World Congress on* (pp. 97-104). IEEE.
- [11] Belle, A., Thiagarajan, R., Sorousmehr, S. M., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. *BioMed research international*, 2015.
- [12] Schmidt, B., & Hildebrandt, A. (2017). Next-generation sequencing: big data meets high performance computing. *Drug discovery today*, 22(4), 712-717.
- [13] Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., & Viboud, C. (2016). Big data for infectious disease surveillance and modeling. *The Journal of infectious diseases*, 214(suppl\_4), S375-S379.
- [14] 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.
- [15] Capobianco, E. (2017). Systems and precision medicine approaches to diabetes heterogeneity: a Big Data perspective. *Clinical and translational medicine*, 6(1), 23.
- [16] Eswari, T., Sampath, P., & Lavanya, S. (2015). Predictive methodology for diabetic data analysis in big data. *Procedia Computer Science*, 50, 203-208.
- [17] Nair, S. S. K., & Ganesh, N. (2016, March). An exploratory study on big data processing: A case study from a biomedical informatics. In *Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on* (pp. 1-4). IEEE.
- [18] Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, 55(1), 64-79.
- [19] Sonnati, R. (2015). Improving Healthcare Using Big Data Analytics. *International Journal of Scientific & Technology Research*, 4(8), 142-146.
- [20] Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2017). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, 11(1), 88-95.
- [21] Al Hamid, H. A., Rahman, S. M. M., Hossain, M. S., Almogren, A., & Alamri, A. (2017). A Security Model for Preserving the Privacy of Medical Big Data in a Healthcare Cloud Using a Fog Computing Facility with Pairing-Based Cryptography. *IEEE Access*, 5, 22313-22328.
- [22] You, I., Choo, K. K. R., & Ho, C. L. (2017). A smartphone-based wearable sensors for monitoring real-time physiological data. *Computers & Electrical Engineering*.
- [23] Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences*, 367, 747-765.
- [24] Fan, W., Zhao, D., & Wang, S. (2015, November). A fast statistics and analysis solution of medical service big data. In *Information Technology in Medicine and Education (ITME), 2015 7th International Conference on* (pp. 9-12). IEEE.
- [25] Rahman, F., Slepian, M., & Mitra, A. (2016, December). A novel big-data processing framework for healthcare applications: Big-data-healthcare-in-a-box. In *Big Data (Big Data), 2016 IEEE International Conference on* (pp. 3548-3555). IEEE.
- [26] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.
- [27] Chrimes, D., Moa, B., Zamani, H., & Kuo, M. H. (2016, August). Interactive Healthcare Big Data Analytics Platform under Simulated Performance. In *Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2016 IEEE 14th Intl C* (pp. 811-818). IEEE.
- [28] Wimmer, H., Yoon, V. Y., & Sugumaran, V. (2016). A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy. *Decision Support Systems*, 88, 51-66.
- [29] Ahmed, E., Yaqoob, I., Hashem, I. A. T., Shuja, J., Imran, M., Guizani, N., & Bakhsh, S. T. (2018). Recent Advances and Challenges in Mobile Big Data. *IEEE Communications Magazine*, 56(2), 102-108.
- [30] Cortés, R., Bonnaire, X., Marin, O., & Sens, P. (2015). Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective. *Procedia Computer Science*, 52, 1004-1009.
- [31] Weider, D. Y., Gottumukkala, C., Senthaiselvi, D. A., Maniraj, P., & Khonde, T. (2016, June). Big Data Analytics in Service Computing: HealthCare-Software As A Service. In *Mobile Services (MS), 2016 IEEE International Conference on* (pp. 170-173). IEEE.
- [32] Baseman, J. G., Revere, D., & Painter, I. (2017, November). Big Data in the Era of Health Information Exchanges: Challenges and Opportunities for Public Health. In *Informatics* (Vol. 4, No. 4, p. 39). Multidisciplinary Digital Publishing Institute.