Big Data Research 4 (2016) 44–58

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

# Towards a Comprehensive Data Analytics Framework for Smart Healthcare Services

Sherif Sakr [a,b,*], Amal Elgammal [c]

[a] King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia
[b] University of New South Wales, Australia
[c] Cairo University, Egypt

ARTICLE INFO

*Article history:*
Received 9 November 2015
Received in revised form 12 April 2016
Accepted 4 May 2016
Available online 3 June 2016

ABSTRACT

With the increasing volumes of information gathered via patient monitoring systems, physicians have been put on increasing pressure for making sophisticated analytical decisions that exploit the various types of data that is being gathered per patient. This phenomenon of continuously growing datasets is arising and gaining momentum in several application domains to what is now recognized in the business community as the *Big Data* challenge. In this article, we define and discuss some of the major challenges in the healthcare systems which can be effectively tackled by the recent advancement in ICT technologies. In particular, we focus on sensing technologies, cloud of computing, internet-of-things and big data analytics systems as emerging technologies which are made possible by the remarkable progress in various aspects including network communication speed, computational capabilities and data storage capacities that provide various advantages and characteristics that can contribute towards improving the efficiency and effectiveness of healthcare services. In addition, we describe the architectural components of our proposed framework, *SmartHealth*, for big data analytics services and describe its various applications in the healthcare domain.



## 1. Introduction

In the last decade, we have been witnessing a continuous increase in lifestyle-related illnesses as a result of various factors such lack of exercise, poor diet, pollution and the addictive habit of smoking. For example, a recent study has estimated that 9% of deaths across the world in 2008 were caused by physical inactivity and merely improving the physical activity levels can affect the life expectancy of the world's population by an increase of 0.68 years [23]. In addition, current guidelines for a healthy life are recommending that adults should spend about 150 minutes of their week on physical activity. However, the study reported that 1/3 of adults are not performing sufficient physical activity which leads to increasing risk of the adults developing certain diseases such as heart-related diseases and diabetes. Our diets have also profoundly differed in the last decades. For example, the consumption of fast food and processed food has been rising continuously across the world resulting in the growing intake of salt, fat, simple sugars and sweeteners [41]. In addition, there has been major increases in the consumption of meat and a corresponding reduction in the consumption of vegetables, whole-grain foods and non-citrus fruits. All of such updates together lead to major increases in the number of calories being consumed, increasing the obesity levels and resulting in severe health threats. Therefore, diseases such as cardiovascular diseases, cancer and diabetes have become the major cause of death across the world.

In general, the medical advancements, improved handling of communicable diseases, and better diet have increased peoples's life expectancy. For example, some reports estimated that life expectancy has increase by 13 years during the course of the 21st century.[1] The UN expects that, life expectancy is going to increase from 68 years in 2005–2010 to 81 years in 2095–2100.[2] Therefore, global aging and its associated effect on the performance of health services are acknowledged as an increasing phenomenon over the last decades. Many countries have been affected by the challenge of an aging population where older individuals represent a bigger share of the size of their population. In particular, the number of people with an age of 65 years old or older is expected to increase from an estimation of 524 million in 2010 to about 1.5 billion in 2050 [13]. Such demographic changes are consequently leading to

* Corresponding author.
*E-mail addresses:* ssakr@cse.unsw.edu.au (S. Sakr), a.elgammal@fci-cu.edu.eg (A. Elgammal).

[1] http://www.bloomberg.com/news/articles/2011-07-21/business-class-longer-lives-and-lower-health-costs.
[2] http://esa.un.org/unpd/wpp/Other-Information/Press_Release_WPP2010.pdf.

http://dx.doi.org/10.1016/j.bdr.2016.05.002

rising demands for healthcare services and higher government expenditures because it is commonly expected that elder individual are naturally more vulnerable to health problems and chronic diseases. Such increasing expenses on healthcare services represent a crucial challenge for almost every government.

In principle, the economics of healthcare systems is getting a lot of attention because of the current dynamics of global demographics. In particular, the spending on healthcare services is commonly a high priority of the internal political agenda and discussions in almost all countries. It is expected that the healthcare costs will account for 20–30% of GDP in some countries by 2050 [17,21], a percentage that is financially unsustainable. In principle, costs continue to increase leading to a consequent need to change the focus of the approaches of healthcare services from that of a reactive model to a model that utilizes predictive healthcare mechanisms. Governments consider smarter healthcare as an effective way of improving quality while minimizing service cost. Establishing such models requires monitoring and diagnosing various data sources for the sake of being able to achieve accurate and effective predictions. Additionally, applications for providing the support of delivering in-home exercise programs for improving the balance and strength of older adults can serve as a preventative action for health problems such as falls. Other applications include activity monitoring, fitness measurements via important signs for monitoring and calorie-intake tracking where various product offerings overlap with main public health advices on activity, exercise, managing one's diet, and observing for early indications of health-related problems. Such indications will be significantly amplified as governments suffer from increasing healthcare expenses in the long run, increasing the opportunities for incorporating sensing-based healthcare platforms and solution.

In principle, the above challenges confirm that the current reactive models of healthcare systems have become unsustainable. Therefore, there have been increasing calls for various major changes in the mechanisms of providing the healthcare services. For example, the healthcare services need to be *predictive* and *proactive* to limit the occurrence of expensive acute health episodes [37]. In addition, the healthcare services need to be individualized, rather than population-based in order to guarantee the delivery of the right treatment. Furthermore, the delivery process of care services need to be decentralized from hospitals to the community and the home. In practice, Information and Communication Technologies (ICT) can play a main role in achieving all of these goals and can represent an effective solution to deliver manageable models of patient services in home and community locations. For example, sensing technology [41] can play a main role on monitoring the main health status indicators of an individual directly or indirectly via ambient monitoring of day-to-day patterns where in-house healthcare has become a main component of the Internet-of-Things (IOT) [30]. In addition, wearable sensing technology [49] is designed to monitor an individual's vital signs at all times, i.e 24/7, where alerts can be communicated to a medical staff/caregiver in the case that a certain limit is reached or in the case an abnormal event such as the collapse of a patient gets observed. Big data analytics services [57] can monitor and detect vital signs and other various measurements which can be provided to physicians or healthcare providers to support them in the diagnosis process. In addition, this diagnosis can be even automated so that it can minimize or remove the need to visit the physician for simple illnesses such as flu and other more common illnesses.

This article focuses on analyzing how the recent advancements of ICT can be effectively exploited and integrated for tackling the above mentioned challenges and contribute towards the state-of-the-art of healthcare services. In particular, we focus on exploiting the advancement in the areas of sensor technologies, cloud computing, Internet-Of-Things and big data analytics systems as emerging technologies, which are made possible by the remarkable progress in various complementary advancements including network communication speed, computational capabilities and data storage capacities that provide various advantages and characteristics that can contribute towards improving the efficiency and effectiveness of healthcare services.

The main contribution of this paper is two-fold; Firstly, analyzing the state-of-the-art in the aforementioned key enabling areas and technologies, and identifying the challenges and gaps for the realization of an integrated and comprehensive smart healthcare data analytics solution. Secondly, drawing on the findings of this analytical study, we propose an integrated and comprehensive framework for big data analytics services in smart healthcare networks, *SmartHealth*, which addresses the revealed challenges and fills in the identified gaps. The Framework acts as a roadmap for the research in the area of big data analytics in smart healthcare applications.

The rest of this paper is organized as follows: the analytical study starts in Section 2 by presenting and analyzing related work and open challenges in the area of smart healthcare systems, which reveals a set of enabling technologies towards the realization of an integrated and comprehensive data analytics smart healthcare solution. We then dedicated separate Sections to discuss and analyze each of these enabling technologies. In particular, sensing technologies is explicated in Section 3, cloud computing and its application to the domain of healthcare is discussed in Section 4. This is followed by a discussion about Big Data storage and processing systems in Section 5. Based on this comprehensive analytical study, an integrated and comprehensive framework for big data analytics in healthcare is derived in Section 6, which represents the second main contribution of this paper. Sample application scenarios and use cases of the proposed framework is presented in Section 7. The paper is then concluded in Section 8, highlighting future work directions.

## 2. Related work and open challenges

The term *Cyber Physical System* represents the umbrella term that integrates and exploits recent ICT advancement in sensing computing, cloud computing, Internet-Of-Things, and big data storage and analytics, with numerous applications in various domains, such as healthcare, manufacturing, traffic, logistics, etc. A cyber-physical system (CPS) is a system of collaborating computational elements controlling physical entities [58]. It is a new revolution in sensing, computing and communication that brings together a variety of resources ranging from networked embedded computers and mobile devices to multimodal data sources including sensor and social; multiple domains such as medical, geographical, environmental, traffic, and behavioral; and diverse situations and application areas such as system health monitoring, chronic medical condition management, disaster response and threat assessment [58]. CPS was identified as a key research area in 2008 by the US National Science Foundation (NSF)[3] and was listed as the number one research priority by the US President's Council of Advisors on Science and Technology [66].

A full-fledged CPS is typically designed as a network of interacting elements comprising physical input and output, as opposed to embedded systems which are typically a set of standalone devices [33]. Ongoing advances in science and engineering improves the link between computational and physical elements, dramatically increasing the adaptability, autonomy, efficiency, functionality, reliability, safety, and usability of cyber-physical systems [33]. This is anticipated to lead to a significant increase in the application

[3] http://www.nsf.gov/.

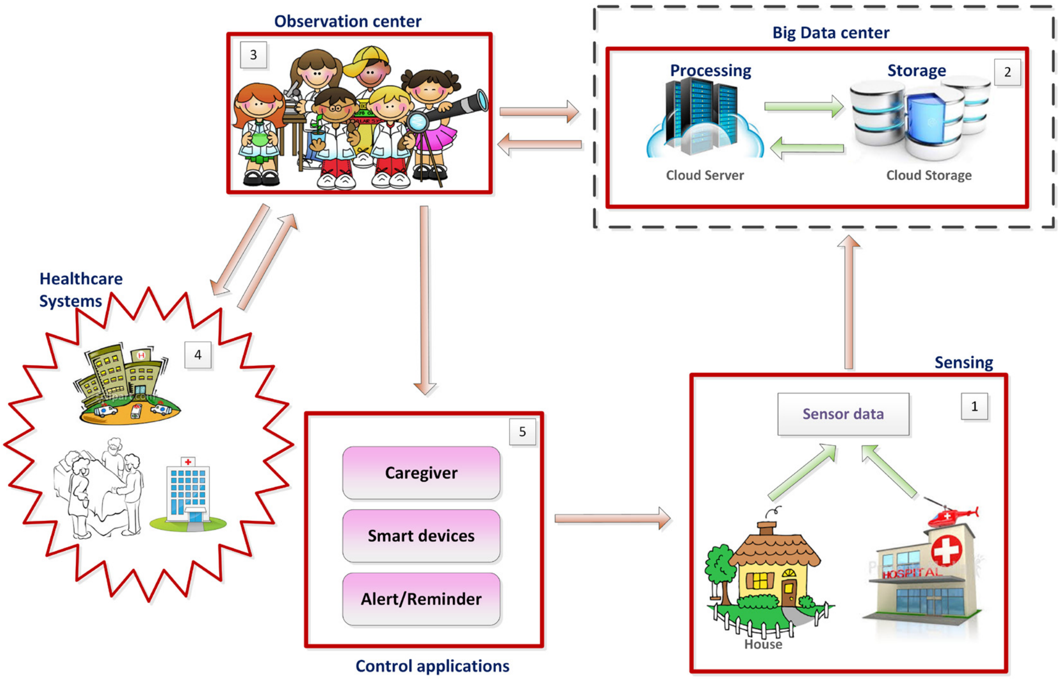


**Fig. 1.** A CPS for healthcare monitoring.

of cyber-physical systems in several dimensions, including: intervention (e.g., collision avoidance), precision (e.g., robotic surgery, nano-level manufacturing), operation in dangerous or inaccessible environments (e.g. search and rescue, firefighting, deep-sea exploration), coordination (e.g., air traffic control, war fighting), efficiency (e.g., zero-net energy buildings) and augmentation of human capabilities (e.g., healthcare monitoring and delivery).

Fig. 1 presents the various interacting components of a typical CPS scenario in healthcare as identified in the literature [22,34,67]. Each component is grounded on one or more enabling technology of CPS (i.e., sensing technologies, IoT, Cloud computing, Big data storage and processing). The main components of the scenario are as follows [22]:

1. *Sensing Component*: This component comprises the utilization of various fixed and wearable sensing devices inside the patient's ambient; i.e., healthcare institution, hospital, home. The sensors continuously monitor and collect a massive amount of data representing the patient's vital signs, which is transmitted to a dedicated data center. This will enable query processing in real time and various big data analytics.
2. *Big Data Center*: This is a data center that is responsible for the storage and processing of data from various sensors by leveraging cloud computing technologies (cf. Section 4). This component represents the main focus of the healthcare framework that will be proposed next in Section 6.
3. *Observation Center*: Clinicians access and make enquiries on patient data in the observation center. Based on the types of data analytics and queries, alarms may be generated and sent to the observation Centre (if necessary). This will enable the observation center team to make informed decisions, some of which are automated.
4. *Healthcare Systems*: Depending on the types of generated alarms, clinicians in the observation center may decide to approach other healthcare systems for consultation. Other healthcare systems will then respond to the clinicians' request.
5. *Control Applications*: Depending on the provided information (from the big data center component) combined with the clinicians' and specialists' medical knowledge (from the healthcare systems component), the clinician in the observation center may decide to send some decisions to the control/actuation component, which could involve invoking an alert/reminder, informing a care giver, or configuring a smart device. Finally, necessary measures will be conducted on the patients based on the invoked control applications.

Only a few CPS applications in healthcare have been proposed to date [24,71] and they lack the flexibility of technology and data integration [22]. Although many CPS architectures have been proposed in the literature, only a few architectures have been proposed that directly address the issues faced by healthcare: [8,35,67]. Arami et al. [35] proposed a service-oriented architecture (SOA)-based medical CPS that, however, lacks a complete structural framework [22]. *CPeSC3* [67] is proposed as a secured CPS architecture for healthcare, which uses WSN-cloud integrated framework. Banerjee et al. [8] proposed an approach to model and analyze medical CPS, however, it also lacks a complete structural framework, such as issues related to security and privacy.

There are some notable CPS applications [7,16,44,65,76] in healthcare. Electronic Medical Records (EMR) [44] constitutes the design of a cyber-physical interface for taking automated vital sign readings. This approach is a solution for vital sign reading which is usually error prone and time consuming. This is a design of a cyber-physical interface that integrate sensors over a wired network that allows retrieving and storing information into an EMR system as structured data. CYPSec is proposed in [65] which is a CPS with environmentally coupled security solutions, that operate by combining traditional security primitives along with environmental features. In [76], Cyber-Physical WBAN System Using Social Networks is proposed as a power game-based approach to mitigate the communication interferences for WBANs based on the people's social interaction information. Medical CPS (MCPS) and Big Data Platform is presented in [16], which proposes a Big Data processing framework for MCPS, which combines the physical world with dynamic provisional, fully elastic cyber world for decision making system in healthcare. Smart Checklist is proposed in [7] to support and guide human participants with their tasks. Interaction with the devices and software applications are also supported by the system. Smart Checklist is expected to assist the medical staff in intensive care to prepare medication, data collection, and other routine activities for patients.

From a data management perspective, the study in [32] proposed a novel information centric approach, such that network-enabled real-time embedded databases, have the capability to handle raw data, and communicate with each other, and control and communicate with wireless sensors in a secure and timely manner. Data merging and integration is addressed in [27] by proposing a data integration process that is divided into two elements: combined and individual. In combined data integration, data collected from multiple sensors can be integrated for further processing. The wide range of data from individual sensors are collected and integrated in individual data integration.

In [67], sensed data involving human health data or human activity detection data was used in this study. A social network based interference mitigation sensing for wireless body area network (WBAN) is proposed in [76], where the ability of mobile phones having speakers and microphones is utilized to send and receive acoustic waves. Acoustic signal processing techniques along with Bluetooth technology were used to measure the physical distance among the mobile phones which acted as a gateway of WBANs. Wiki-Health [73] is a big data platform for health sensor data management. Wiki-Health utilizes cloud computing for sensor data management and processing with a query and analysis layer. The work we propose in this project is very close to Wiki-Health, however, we intend to target and semantically integrate multimodal data from heterogeneous sensors, patient's information, contextual information, HER and health documents. Multimodal patient's information will then be integrated to the patient treatment process, which will enable physicians to make more informed decisions.

Medical status monitoring applications include [26,29,53,59,70]. MobiHealth [29] represents the effort to gather data from the wearable sensor devices that people carry all day. This project is one of the early efforts made to monitor medical statuses from sensors. It tries to collect audio and video signal to provide early response in case of accidents. CodeBlue [59] is a platform that is made up of integrated biomedical sensors such as two-lead ECG, pulse oximeter, and motion sensor with pub/sub based routing software architecture. CodeBlue manages and communicates among the medical devices. This system is a pioneering project in terms of its early use of in-network aggregation and smart routing. AlarmNet [70] is a wireless biosensor network system prototype consisting of heart rate, pulse rate, oxygen saturation and ECG. Environmental parameters such as temperature and humidity provide spatial contextual data. Privacy, power management, and query management are also considered in the system. Similarly, Mobile ECG [26] system uses smart mobile phones as the base station for ECG measurement and analysis. In Predicting Vital Signs [53] a system has been developed that can predict the heart rate, blood pressure, and other vital signs around 20 seconds earlier than it was possible before.

*Cyber physical systems enabling technologies* is enabled by the interaction and interplay between a number of key enabling technologies. More specifically, a primary key enabler is the advancement in *sensing technologies* ('sensing component' in Fig. 1). A tightly related term in the context of healthcare monitoring is *participatory sensing* (also known as crowd sensing, urban sensing, human-centered sensing, or opportunistic sensing), which refers to the phenomenon of communities (or other groups of people) contributing sensory information to form a body of knowledge [20]. With the advent of smartphones and various low-cost wearing devices and fabrics, many citizens are actively carrying sensors on them for long durations of time. In the healthcare domain, patients can actively participate in their daily health monitoring, which provide them with a means to live independently, preserve their privacy and ultimately improve their quality of life. As pointed out by [61], "Health and fitness" is one of the major promising applications of participatory sensing in practice. Sensing technologies will be discussed in more detail next in Section 3.

*Internet of Things* (IoT) is another enabler technology of CPSs (cf. the link between 'Sensing' and 'Big Data Center' components in Fig. 1). IoT is defined as "*the interconnection of uniquely identifiable computing devices via Internet connectivity*" [30]. In practice, IoT is projected to provide rich connectivity among devices, services and systems. The connectivity of these objects is expected to enable automation in many areas of application [6]. *Things*, in the context of healthcare systems, refer to a wide variety of devices such as heart monitoring implants and biochip transponders on farm animals, etc.

*Cloud computing* is another major player in the success of CPS, which allows computing services/resources to be provided as utilities over a network, usually over the internet [61]. Cloud computing technology is utilized to store and process gathered sensed data as shown in the 'Big Data Center' component of Fig. 1. Cloud computing and its role in enabling CPS with a special focus on healthcare will be covered in Section 4. In addition to the glut of new and novel application areas anticipated, IoT is also expected to generate large amounts of data from *diverse locations* and from *heterogeneous sources*, which raises many challenges that will be discussed in more detail in Section 5 of this article.

With such a wide spectrum of systems with various design decisions, it becomes difficult for the systems designers and developers to decide which system to use to meet the requirements for the application at hand. The part in Fig. 1 that is highlighted by dashed lines, i.e., 'Big Data Center', represents the main focus of the framework we propose in Section 6. This includes the pre-processing, storage, maintenance, semantic integration of heterogeneous multi-modal sensor data (big data), and their efficient processing and analysis supported by cloud computing technology. Our comprehensive data analytics framework is derived based on this comprehensive analytical study, and aims to addressing the challenges and filling in the gaps identified throughout this paper.

## 3. Sensing technologies

Sensing is a pervasively used technology in nearly every aspect of hospital-based service starting from the simplest digital thermometer to complex laser-guided surgical tools [41]. For example, imaging sensing technologies (e.g., magnetic, X-ray), positron emission tomography (PET), computed tomography (CT) and ultrasound are commonly used technologies for providing the medical staff with several insights into the health status of every patient. These sensors have played a crucial role in transforming diagnostic medicine. In particular, such information enable physicians to identify areas of injury or abnormality, conduct minimally invasive surgery, and assess the success or failure of a medical operation. For instance, in obstetric care, ultrasound enables the physician to track and investigate the developing fetus and determine any fetal or other abnormalities that may affect the health of the mother or baby. Furthermore, advanced sensing devices are utilized by pathologists in hospital laboratories to perform hematology, immunology, biochemistry, histopathology and microbiology functions. Sensors can also have a significant role in the medical treatment process. For instance, they are able to detect situations such as missed heartbeats. In addition, they can play a main role in optimizing the drug delivery process through determining the right time to provide a drug. Wearable, disposable vital sign sensors are starting to emerge providing low-cost, continuous monitoring of vital signs for patients, regardless of location or health status. It is estimated that by 2018, 5 million of such type of sensors will be

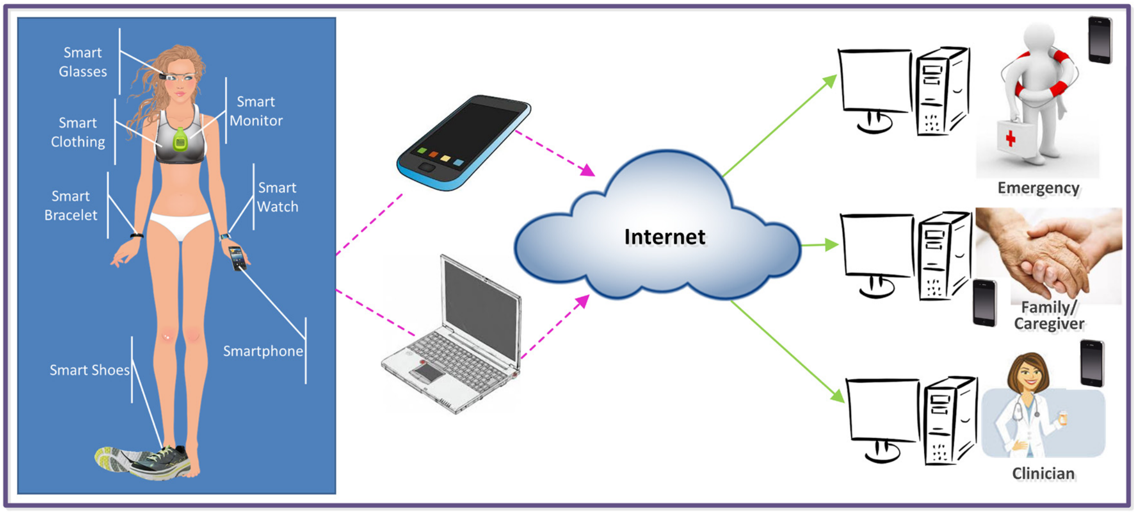


**Fig. 2.** A typical architecture of wearable sensors connected via the cloud.

sold.[4] For example, the *LifeShirt* System[5] is an example of a multi-sensor continuous monitoring system for collecting, analyzing and reporting a patient's health data. It captures the ongoing patient's data rather than data snapshots which is usually collected during the patient's periodic visits to the doctor's office. The left-hand side of Fig. 2 shows different examples of these wearable devices, e.g., smart shoes, smart watch, smart bracelet, etc., continuous monitoring of vital signs.

In general, a *sensor* is defined as a device that converts a physical measure into a signal that is read by an observer or by an instrument [64]. In principle, sensors exploit a wide spectrum of transducer and signal transformation approaches with corresponding variations in technical complexity. These range from relatively simple temperature measurement based on a bimetallic thermocouple, to the detection of specific bacteria species using sophisticated optical systems. In the context of the healthcare and wellness, there are various sensing approaches such as optical sensors, mechanical sensors, microelectromechanical systems (MEMS), electrochemical, semiconductor and biosensing [41]. In practice, sensor technologies have various applications in the healthcare domain such as facilitating the process of physiological monitoring (e.g., blood pressure, heart rate) and screening applications (e.g., blood biochemistry and falls risk estimation). In addition, in domains such as home and community health, telemonitoring, telehealth and mobile health sensor applications enable remote monitoring and management of patients with chronic diseases such as diabetes and congestive heart failure (CHF). Furthermore, in hospitals and primary healthcare facilities, the sensor applications are more focused on medical screening and diagnostics applications such as electrolyte-level measurement, point-of-care blood chemistry testing and analyzing blood gas concentrations. There is an increasing growth in the market of diagnostic sensors that perform pregnancy testing, cholesterol monitoring, food allergy testing and DNA testing. In particular, examples of key application areas of sensors in clinical healthcare are [41]:

– *Imaging*: Low cost CCD and ultrasound sensors are used for medical imaging [45]. Smart pills can be used for intestinal imaging [40].
– *Screening and Diagnostics*: Optical and biochemical sensors are used for diagnostics applications and point-of-care monitoring [72]. In addition, biosensors can be used to determine drugs, proteins levels, bacterial infection and hormones in biological samples [42].
– *Motion and Kinematics*: Accelerometer and gyroscopes are examples of body-worn wireless sensors which can be used to determine balance and the fall risk aspect and to track the impact of medical interventions. Kinematic sensors are used in the evaluation process of prosthetic limb replacements [4]. They are also utilized in stroke rehabilitation to track the status of identified physical exercises [62].
– *Physiological*: Main physiological indicators of health status such as blood pressure and ECG/EKG can be measured and tracked [39].

In general, with the increasing awareness of people about their health, there is an associated growth in the market for various types of sensors which can be used to monitor and track the progress in wellness programs (e.g., obesity prevention) [12]. Sensors such as blood-pressure monitors, body-worn heart-rate and integrated activity monitors, and pulse oximeters are increasingly being used in this growing application domain. Another key element of wellness where sensors can have effective applications is personal safety especially in the home environments. For example, the use of smoke detectors have been widely used and long established [46]. Semiconductor or electrochemical sensors are commonly used in residential carbon monoxide sensors. Furthermore, the wide availability and usage of general positioning tracking sensors (GPS) (e.g., accelerometers in smartphones) are usually used for recreational purposes to enable real-time pace, elevation, location and direction information to be used by a jogger or personal safety applications. For example, the GPS sensors can be used to track and identify the location of a child or older adults.

In practice, the technology of ubiquitous sensing relies on distributed and networked sensors for monitoring user activities while remaining transparent to the users [48]. For instance, for behavioral monitoring, RFID and wireless sensors are used to detect interactions between humans and their environment. In principle, sensor systems are designed to enrich the functions of smart sensors by supporting more features including communications (wired and wireless), data collection and preparation, display, enclosures and mounts, security and remote manageability. The requirements mix of system capabilities is usually defined by the application

[4] http://mobihealthnews.com/22089/2018-5-million-disposable-mobile-medical-sensors/.
[5] http://vivonoetics.com/products/sensors/lifeshirt/.

needs. For example, modern smartphones and tablets is equipped with various integrated sensors which are employed by the operating system to improve the user experience [19]. In addition, such devices usually include location and motion location sensors (e.g., pressure sensors, gyroscopes, accelerometers), optical sensors (ambient light sensors, image sensors, proximity sensors, and display sensors), silicon microphones, and various other environment sensors. The sensors are integrated into one of the most frequently used devices which guarantee high user compliance. In addition, currently, the most popular mobile operating systems (e.g., Android, iOS, and Windows 8) provide sensor-based development frameworks (e.g. *Run Keeper*[6]) that enable programmers to easily access the data streams of these sensors in their applications. Such integration of sensor frameworks into the process of software development highly facilitate the application development process especially as the readings from a single sensor cannot always be used to provide the required comprehensive measurements for deciding about the required target actions. In various scenarios, the measurements from various sensors might be needed to either fully understand the measurement of interest or understand the context sensitivity for situational awareness of the measurement. For example, non-contact measurements of gait velocity would need to use multiple sensors at fixed distances to be able to calculate the velocity of individuals as they move past. Motion analysis applications (e.g., Wii and Kinect) have utilized 3D-gyroscope, 3D-accelerometer and 3D-magnetometers in their solutions. The mechanisms of utilizing and fusing sensor data are based on the application's needs and available computing resources.

We believe that sensors technologies can take a main role in developing healthcare services with smart software capabilities that provide a layer of data analysis and interpretation to provide physicians and healthcare staff in their work. In practice, current technology advances have enabled progressively speedy and increasing exploitation of sensing technologies. For example, as of this writing, more than one billion smartphones have been sold.[7] In 2013, for the first time, the sales of smartphone has exceeded the sales of traditional mobile devices.[8] 3G mobile broadband connectivity has become more widely accessible with the faster 4G broadband connectivity being introduced. In principle, connectivity, whether via 3G, 4G, General Packet Radio Service (GPRS), Wi-Fi or Bluetooth, have become pervasive. Furthermore, cloud-enabled services have been increasingly utilized to provide ever-increasing data sharing, storage, processing, and aggregation capabilities. Therefore, comprehensive integration between sensing technologies and ICT technologies will provide patients with more opportunities to produce information about their health status and to be engaged in the process of managing their own status with their physicians. In addition, it will provide them with the capability to exploit and control the information in a way that has not been possible before. In practice, the data from these various sensing devices will provide physicians with more information than they can manually analyze.

Therefore, physicians need to be equipped with the right services that support them to achieve a patient care model which is based on predictive, proactive, preventive, and personalized medications, so that they can effectively and accurately determine the state of their patients and prescribe an appropriate course of treatment.

6 http://runkeeper.com.

7 http://news.cnet.com/8301-1035_3-57534132-94/worldwide-smartphone-user-base-hits-1-billion/.

8 http://www.3news.co.nz/technology/smartphones-now-outsell-dumb-phones-2013042912.

## 4. Cloud computing

Cloud computing has been acknowledged on the top of Gartner's list of the ten most disruptive technologies of the next years [1]. It represents a paradigm shift in the field of ICT which has been already emerging to change the ways of how businesses deal with their storage and computing resources [18]. In principle, cloud computing represents an emerging paradigm for the process of provisioning computing resources and infrastructure. This paradigm shifts the location of the computing resources and infrastructure to the network with the goal of reducing the costs of managing the software and hardware resources. As a result, users and businesses are able to pervasively access application services from anywhere in the world remotely and on demand. Therefore, cloud computing has been considered as a significant step towards achieving the long-held dream of envisioning computing as a utility [5] where the economies of scale principles help to drive the cost of computing infrastructure effectively down. In practice, big players of the technology companies (e.g., Amazon, Microsoft, Google, IBM) have been quite active on establishing their own data centers across the world to ensure reliability by providing redundancy for their provided infrastructure, platforms and applications to the end users. Therefore, cloud-based services offer several advantages such as: flexibility and scalability of storage, computing and application resources, optimal utilization of infrastructure and reduced costs. Therefore, cloud computing provides a great chance to supply storage, processing, and visualization resources for sensing-based healthcare data. In particular, healthcare sensors can be discrete or part of a geographically distributed network. This cloud-based resources are capable of accommodating dynamic workloads.

In fact, the discussion in industry and academia has taken a while until they are able to define the road map for defining what cloud computing actually means [63,51,60]. The US National Institute of Standards and Technology (NIST) has published a definition that reflects the most commonly agreed features of cloud computing. This definition describes the cloud computing technology as: "*a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*". In principle, one of the important features provided by the cloud computing technology is that computing hardware and software capabilities are made accessible via the network and accessed through standard mechanisms that can be supported by heterogeneous thin or fat client platforms (e.g., laptops, mobile phones, and PDAs).

In cloud computing, the provider's computing resources are pooled to serve multiple consumers using a multitenant model with various virtual and physical resources dynamically assigned and reassigned based on the demand of the application workload. Therefore, it achieves the sense of location independence. Examples of such shared computing resources include storage, memory, network bandwidth, processing, virtual networks and virtual machines. In practice, one of the main principles for the data centers technology is to exploit the virtualization technology to increase the utilization of computing resources. Hence, it supplies the main ingredients of computing resources such as CPUs, storage, and network bandwidth as a commodity at low unit cost. Therefore, users of cloud services do not need to be concerned about the problem of resource scalability because the provided resources can be virtually considered as being infinite. In particular, the business model of public cloud providers rely on the mass-acquisition of IT resources which are made available to cloud consumers via various attractive pricing models and leasing packages. This provides applications or enterprises with the opportunity to gain access to powerful infrastructure without the need to purchase it.

In practice, pooled IT resources that are made available and shared by multiple cloud users leads to increasing or even maximizing the utilization of the shared resources. Operational costs and inefficiencies can also be significantly reduced by relying on best practices and patterns for optimizing cloud architectures, their governance, and their management of cloud resources. In addition, cloud platforms are equipped with the tools and technologies which are designed to dynamically and instantly allocate computing resources to cloud users on-demand and according to the application demand. Such feature enables cloud users to scale the computing resources in order to elastically handle the peaks and fluctuations in their workloads. Such feature increases the scale of the financial gain that can be gained from relying on cloud resources by automating the down scaling, the capability of computing resources to continuously fulfill unpredictable usage demands and avoiding potential loss or degradation of application request that can happen when handling fluctuated workloads. The availability and reliability of the cloud computing resources is another significant factor which is directly contributing to the tangible business benefits. In particular, a hallmark of the typical cloud platform is its guaranteed ability to provide extensive support for increasing the availability of a cloud-based computing resource in order to minimize or even eliminate outages, and for increasing its reliability so as to minimize the impact of runtime failure conditions [3]. For example, *Microsoft HealthVault*[9] has been introduced as a web-based platform for storing and maintaining health and fitness information.

In particular, the platform connects with a massive array of medical devices, consumer gadgets and applications to automatically import data. In addition, the user can manually add data, including weight, lab results, medications, menstruation dates, immunizations, medical procedures, and more. Furthermore, doctors can send data and files of their patients right into their HealthVault account. Thus, HealthVault provides the service of implementing an online and widely accessible personal medical record. *Sensor-Cloud* [74] has been introduced as an infrastructure with the aims of managing physical sensors by connecting them to the cloud. In particular, Sensor-Cloud bundles the physical sensors into virtual sensors where users can combine them together to achieve advanced results. *SenaaS* (Sensor-as-a-Service) [2] is another approach that has been proposed to encapsulate both physical and virtual sensors into services according to the Service Oriented Architecture (SOA) approach. SenaaS mainly focuses on providing sensor management as a service rather than providing sensor data (collection and dissemination) as a service.

In practice, we believe that the integration of sensing-based technologies into various aspects of our daily activities will produce massive amounts of data which is going to only increase as progressively more sensor data are fed into the data storage systems of the healthcare platforms. In principle, the notion of *Big Data* has acquired increasing interest in recent years as a reflection to the massive amounts of data which is generated each day using the recent advancements in ICT [55]. Sensing-based technologies are projected to participate as one of the fundamental contributors to this big data era especially as the Internet of Things (IOT) has started to gain attention in our daily activities [68]. It is also expected that big sensor data is going to heavily leverage the capabilities of cloud computing resources for storing, processing, and visualizing massive amounts of data where the access to the generated data will become more pervasive, specifically via smartphones and small devices. In addition, it facilitates

[9] https://www.healthvault.com/sa/en.

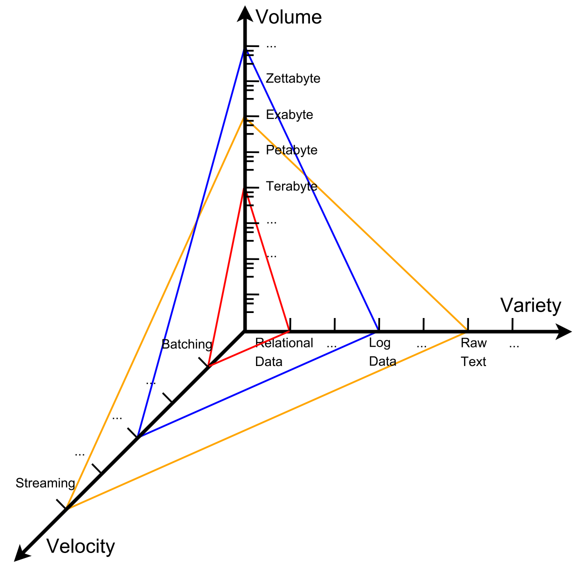


**Fig. 3.** 3V Characteristics of Big Data.

the ability for combining sensor-based healthcare data with other sources of sensor data in innovative mechanisms to discover new insights. Fig. 2 depicts the general architecture of patients/citizens equipped with various wearable (medical) devices, e.g., ECG and respiration, motion sensors, etc., which are connected on the cloud with clinicians, emergency services providers and caregivers. Collected sensors data about patients vital signs along with a patient's information and history can then be abstracted and analyzed to generate meaningful information, which gives more insight to clinicians to make informed decisions and to predict potential threats.

## 5. Big data storage and processing systems

In general, using the mining metaphor, data represent the new gold where analytics systems represent the machinery that mines, molds and mints it. In practice, healthcare systems across the world are facing the challenge of information overload in caring for patients. Healthcare analytics is defined as a set of computer-based methods, processes and workflows for transforming raw health data into meaningful insights, new discovery and knowledge that helps in making more effective healthcare decisions [11]. Recently, healthcare analytics has been gaining increasing importance not just for the sake of improving the patient and healthcare services but it is also considered as an effective tool to reduce costs. In particular, healthcare analytics is now gaining popularity thanks to the eruption of the *Big Data* phenomena in the healthcare domain.

In general, big data is known by its three key 3V characteristics (Fig. 3): *V*olume (refers to the scale of the size of the data), *V*elocity (represents the streaming data and large-volume data movements) and *V*ariety (refers to the complexity of data in many different structures, ranging from relational to logs to raw text) [55]. For example, in healthcare data, the structured data can come in the form of electronic medical records (EMRs) or electronic health records (EHRs) that include familiar input record fields such as patient name, data of birth, address, hospital name, address and other field-based information. Semi-structured data includes instrument readings and data generated by the ongoing conversion of paper records to electronic health and medical records. Furthermore, structured and unstructured data streams can be cascading into the healthcare systems from various sources such as fitness devices, genetics and genomics, social media and other sources. Thus, healthcare organizations are in crucial need for efficient ways to combine and convert such varieties of data including automating the conversion from structured to unstructured data.

In practice, we are now data rich but information poor. Therefore, big data technologies provide a powerful infrastructure for gleaning actionable information from this sea of data. According to a McKinsey report, big data analytics is the platform to deliver five values to healthcare: Right Living, Right Care, Right Provider, Right Value and Right Innovation [28]. These values provide boundless opportunities in improving healthcare services on one hand and reducing waste and costs on the other. For example, healthcare analytics can help with patient classification based not only on simple traditional demographic attributes such as gender, age and life style but also by relevant health and clinical characteristics related to medical conditions, risk propensities, genetic disposition and therapeutic probabilities. Healthcare analytics provide the ability to optimize and tailor the course of care to each individual patient based on multitudes of factors that go into defining the medical protocol of care for such patients: prior medical history, precautions, allergies, genetic traits, personal risk factors, work and life styles and safety management. Healthcare analytics can also uncover causal relationships between a number of quality indicators and factors that influence or affect those health indicators in a patient population or specific patients. Furthermore, healthcare analytics can be used to calculate more accurate measures of patient risk stratification, determining the level of health complication, comorbidity impact and how serious a patient's health status will affect the outcome. From such calculations, risk profiles about patients can be determined which will help in designing care plans for the population of patients with the same profiles. Such classification would be very helpful to any accountable care organizations' plans and pricing. In practice, achieving these goals require the availability of powerful big data storage, processing and analytics platforms.

In the last decade, we have seen continuous advancements in the domain of big data technologies. In the following section, we give an overview of these technologies classified under two main categories: big data storage systems (Section 5.1) and big data processing systems (Section 5.2).

### *5.1. Big data storage systems*

In general, relational database management systems (e.g. MySQL, PostgreSQL, SQL Server, Oracle) have been considered as the *one-size-fits-all* solution for data persistence and retrieval for decades. They have matured after extensive research and development efforts and very successfully created a large market and many solutions in different business domains. However, the ever increasing need for scalability and new application requirements have created new challenges for traditional RDBMS. In particular, currently, we are witnessing a continuous increase of user-driven and user-generated data that resulted in a tremendous growth in the type and volume of data which is produced, stored and analyzed. For example, various newer sets of sources data generation technologies such as: sensor technologies, automated trackers, Global Positioning Systems (GPS) and monitoring devices are producing massive datasets. In addition to the speedy data growth, data has also become increasingly of sparse and semi-structured in nature. Such changes led to the challenge that the traditional data management techniques that required upfront schema definition and relational-based data organization is inadequate in many scenarios. Therefore, in order to tackle this challenge, recently, we have witnessed the emergence of a new generation of scalable data storage system called *NoSQL* (**N**ot **O**nly **SQL**) database systems. This new class of database systems come into four main types:

- *Key-value stores*: These systems use the simplest data model which is a collection of objects where each object has a unique key and a set of attribute/value pairs.
- *Extensible record stores*: They provide variable-width tables (Column Families) that can be partitioned vertically and horizontally across multiple servers.
- *Document stores*: The data model of these systems consists of objects with a variable number of attributes with a possibility of having nested objects.
- *Graph stores*: The data model of these systems uses graph structures with nodes, edges and properties to represent and store data.

In general, scalability represents the capability of a system to increase throughput via increasing the allocated resources to handle the increasing workloads. In practice, scalability is usually accomplished either by provisioning additional resources to meet the increasing demands (vertical scalability) or it can be accomplished by grouping a cluster of commodity machines to act as an integrated work unit (horizontal scalability). In principle, vertical scaling option is typically expensive and proprietary while horizontal scaling is achieved by adding more nodes to manage additional workloads which fits well with the *pay-as-you-go* pricing philosophy of the emerging cloud computing models. In addition, vertical scalability normally faces an absolute limit that cannot be exceeded, no matter how much resources can be added or how much money one can spend. Furthermore, horizontal scalability leads to the fact that the storage system would become more resilient to fluctuations in the workload because handling of separate requests are handled such that they do not have to compete on shared hardware resources.

In practice, while there are many systems[10] that are identified to fall under the umbrella of NoSQL systems are quite varied, however, each of these systems come with their unique sets of features and value propositions [56]. For example, the key/value (KV) data stores represent the simplest model of NoSQL systems which pairs keys to values in a very similar fashion to how a map (or hashtable) works in any standard programming language. Various open source projects have been implemented to provide key-value NoSQL database systems such *Memcached*,[11] *Voldemort*,[12] *Redis*[13] and *Riak*.[14] Columnar, or column-oriented, is another type of NoSQL databases. In such systems, data from a given column is stored together in contrast to a row-oriented database (e.g., relational database systems) which keeps information about each row together. In column-oriented databases, adding new columns is quite flexible and is performed on the fly on a row-by-row basis. In particular, every row may have a different set of columns which allow tables to be sparse without introducing any additional storage cost for null values. In principle, columnar NoSQL systems represent a midway between relational and key-value stores. *Apache HBase*[15] is currently the most popular open source system of this category. Another category of NoSQL systems is document-oriented database stores. In this category, a document is like a hash, with a unique ID field and values that may be any of a variety of types, including more hashes. In particular, documents can contain nested structures, and so they provide a high degree of flexibility, allowing for variable domains. *MongoDB*[16] and *CouchDB*[17] are currently the

10 http://nosql-database.org/.
11 http://memcached.org/.
12 http://www.project-voldemort.com/voldemort/.
13 http://redis.io/.
14 http://basho.com/riak/.
15 http://hbase.apache.org/.
16 http://www.mongodb.org/.
17 http://couchdb.apache.org/.

two most popular systems in this category. Finally, NoSQL Graph databases is another category which excel in handling highly interconnected data. In principle, a graph database consists of nodes and relationships between nodes where both relationships and nodes can be described using descriptive information and properties (key-value pairs). In principle, the main advantage of graph databases is that they provide easy functionalities for traversing through the nodes of the graph structure by following relationships. The *Neo4J*[18] database system is currently the most popular in this category.

### 5.2. Big data processing and analytics systems

In general, data are not useful in and of themselves. They only have utility if meaning and value can be extracted from them. Therefore, given their utility and value, there are always continuous increasing efforts that are devoted towards producing and analyzing them. In principle, big data discovery enables data scientists and other analysts to uncover patterns and correlations through analysis of large volumes of data of diverse types. Insights gleaned from big data discovery can provide businesses with significant competitive advantages, such as more successful marketing campaigns, decreased customer churn, and reduced loss from fraud. In practice, the increasing demand for large-scale data processing and data analysis applications has triggered the development of novel solutions from the industry and the academia. In principle, in the last decade, the MapReduce framework [15] has represented the defacto standard of big data technologies and has been widely utilized as a fundamental mechanism to harness the power of clusters of commodity machines. In general, the fundamental principle of the MapReduce framework is to move analysis to the data, rather than moving the data to a system that can analyze it. In addition, one of the main advantages of the MapReduce framework is that it isolates the application from the details of running a distributed program, such as issues on data distribution, scheduling and fault tolerance. In particular, it allows programmers to think in a *data-centric* fashion where they can focus on applying transformations to sets of data records while the details of distributed execution and fault tolerance are transparently managed by the MapReduce framework. In practice, the Hadoop project,[19] the open source realization of the MapReduce framework, has been a big success and created an increasing momentum in the research and business domains. For example, due to its success, it has been supported by many big players in their big data commercial platforms such as *Microsoft*,[20] *IBM*[21] and *Oracle*.[22] In addition, several successful startups such as *MapR*,[23] *Cloudera*,[24] *Platfora*[25] and *Trifcata*[26] have built their solutions and services based on the Hadoop project. Fig. 4 illustrates Google's web search trends for the two search items: *Big Data* and *Hadoop*, according to the Google trend analysis tool.[27] In principle, Fig. 4 shows that the search item *Hadoop* has overtaken the search item *Big Data* and have since dominated the Web users' search requests during the period between 2008 and 2012 while since 2013, the two search items have started to go side by side.

In general, the discovery process often employs analytics techniques from a variety of genres such as time-series analysis, text

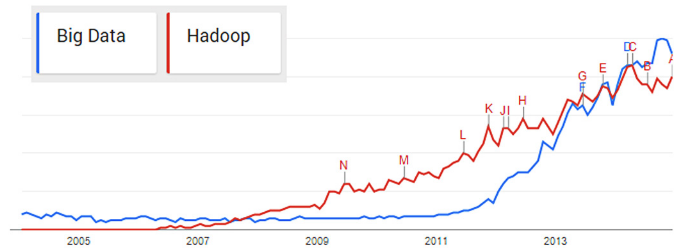


**Fig. 4.** Google's Web search trends for the two search items: Big Data and Hadoop (created by Google trends).

analytics, statistics, and machine learning. Moreover, the process might involve the analysis of structured data from conventional transactional sources, in conjunction with the analysis of multi-structured data from other sources such as click streams, call detail records, application logs or text from call center records. In practice, in the last few years, both the research and industrial communities have recognized some main limitations in the MapReduce/Hadoop framework [57] and it has been acknowledged that it can not be the *one-size-fits-all* solution for all big data processing problems. For example, in processing large-scale structured data, several studies reported on the significant inefficiency of the Hadoop framework. In particular, the studies claim that Hadoop is the wrong choice for interactive queries that have a target response time of a few seconds or milliseconds [50]. In addition, many programmers may be unfamiliar with the Hadoop framework and they would prefer to use SQL as a high level declarative language to implement their jobs while delegating all of the optimization details in the execution process to the underlying engine [57]. As a result, Google has designed the Dremel system [43], which is commercialized through the system of *BigQuery*,[28] to provide interactive analysis of nested data. Other projects which have been designed to tackle these challenges include *Apache HIVE*[29] and *Cloudera Impala*[30] which have been introduced to support the SQL flavor on top of the Hadoop infrastructure and provide competing and scalable performance on processing large scale structured data.

Nowadays, graphs with millions and billions of nodes and edges have become very common. The enormous growth in graph sizes requires huge amounts of computational power to analyze. In general, graph processing algorithms are iterative and need to traverse the graph in a certain way. In practice, graph processing tasks can be implemented as a sequence of chained MapReduce jobs that involves passing the entire state of the graph from one step to the following. However, this mechanism is not suited for graph analysis and leads to inefficient performance because of the communication overhead and associated serialization overhead plus the additional requirement of coordinating the steps of a chained MapReduce [57]. Therefore, in 2010, Google introduced the *Pregel* system [36], open-sourced as *Apache Giraph*,[31] that uses a bulk synchronous parallel (BSP) based programming model for efficient and scalable processing of big graphs using clusters of commodity machines. Several other projects (e.g., *Trinity*[32] and *GraphLab*[33]) were introduced to tackle the problem of large scale graph processing.

In general, stream computing is a new paradigm which has been necessitated by new data-generating scenarios, such as the

18 http://neo4j.com/.
19 http://hadoop.apache.org/.
20 http://azure.microsoft.com/en-us/services/hdinsight/.
21 http://www-01.ibm.com/software/data/infosphere/hadoop/enterprise.html.
22 http://www.oracle.com/us/products/middleware/data-integration/hadoop/overview/index.html.
23 https://www.mapr.com/.
24 http://www.cloudera.com/.
25 https://www.platfora.com/.
26 http://www.trifacta.com/.
27 http://www.google.com/trends/.
28 https://cloud.google.com/bigquery/what-is-bigquery.
29 https://hive.apache.org/.
30 http://www.cloudera.com/content/cloudera/en/products-and-services/cdh/impala.html.
31 https://giraph.apache.org/.
32 http://research.microsoft.com/en-us/projects/trinity/.
33 http://graphlab.com/.

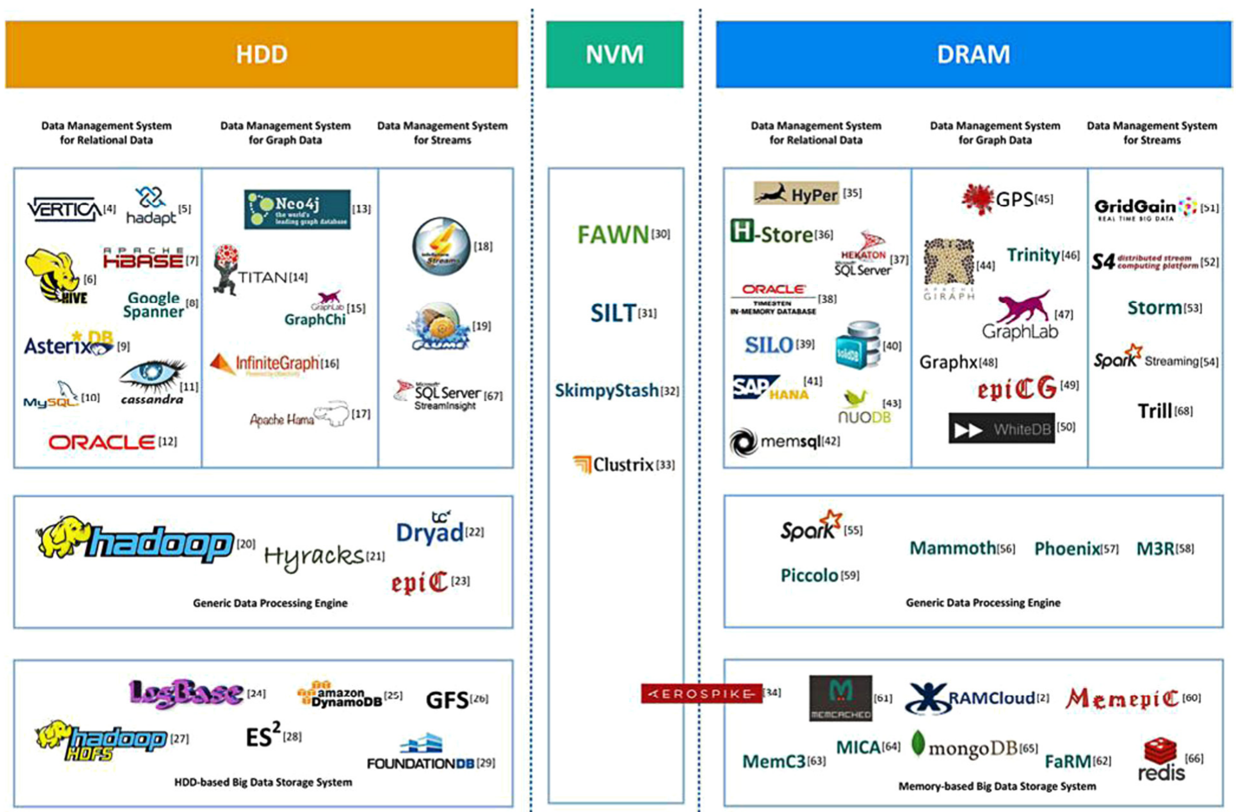


**Fig. 5.** The (partial) landscape of big data systems [75].

ubiquity of mobile devices, location services, and sensor pervasiveness. In general, stream processing engines enables a large class of applications in which data are produced from various sources and are moved asynchronously to processing nodes. Thus, streaming applications are normally configured as continuous tasks in which their execution starts from the time of their inception till the time of their cancellation. In particular, in static data computation, questions are asked of static data. In streaming data computation, data is continuously evaluated by static questions. In principle, the fundamental architecture of the Hadoop framework is designed with the assumption that the whole output of each map and reduce step needs to be stored into the local storage before it can be processed by the next node. This materialization step enables the implementation of a simple and straightforward checkpointing/restarting fault tolerance mechanism. Therefore, in scalable analysis of streaming data, Hadoop had been shown to be an inadequate platform as well [57].

As a result, Twitter has released the Storm system[34] that fills this gap by providing a distributed and fault tolerant platform for implementing continuous and real-time processing applications of streamed data. The main abstraction in the Storm programming model is the *stream* which represents an unbounded sequence of tuples. The basic programming abstractions of Storm that supports the basic functionalities for implementing stream processing operations are *spouts* and *bolts*. A spout represents the source of streams. A bolt can receive an arbitrary number of input streams, perform some processing according to the application logic, and optionally produces new streams. Complex stream processing operations (e.g., the determination of trending topics by analyzing a stream of tweets) would require several steps and consequently multiple bolts. A storm topology is represented as a graph of stream operators where each node is a spout or bolt. Edges in the graph describes which bolts are associated to which streams. Once a spout or bolt sends a tuple to a stream, it sends the tuple to each bolt which is associated to that stream. Links between nodes in a topology illustrates how data tuples should be transferred over the nodes of the cluster. It should be noted that all nodes of a Storm topology are executed in parallel. For any topology, we can determine the degree of parallelism which is needed for each node, and then Storm will create the number of threads across the cluster to perform the execution. Fig. 4 depicts a sample Storm topology. Other systems in this domain include *IBM InfoSphere Streams*[35] and *Apache S4*.[36] In practice, big stream processing systems are very well-suited for dealing with data streams which are generated from various types of sensing devices.

*Apache Mahout*[37] is an open source project which is designed to solve very practical and scalable machine learning problems. In particular, Mahout is essentially a set of Java libraries which is well integrated with Apache Hadoop and is designed to make machine learning applications easier to build. Mahout enables its users to get started with common use cases quickly. Similar to Apache Hive, which provides an SQL-like interface to querying data in Hadoop's data, Mahout translates machine learning tasks expressed in Java into MapReduce jobs. The AMPLab, a group of researchers from UC Berkeley, has developed a project called *MLbase* [31] as general-purpose machine learning library which has a similar goal to Mahout's goal to provide a viable solution for dealing with large-scale machine learning tasks. *R*[38] is an open-source software package that facilitate performing statistical analysis on data. In particular, *R* provides a programming language which is used by statisticians,

34 https://storm.apache.org/.

35 http://www-03.ibm.com/software/products/en/infosphere-streams.

36 http://incubator.apache.org/s4/.

37 https://mahout.apache.org/.

38 http://www.r-project.org/.

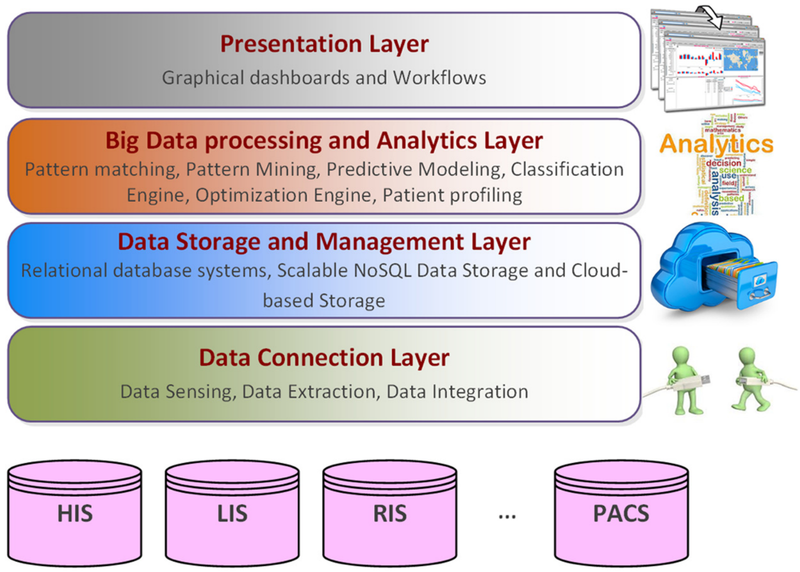


**Fig. 6.** The layered architecture of SmartHealth Framework.

data scientists and others who seek to conduct statistical analysis of data and discover new insights from data using mechanisms, such as clustering, regression, classification and text analysis. *R* supports a rich set of statistical, machine learning and graphical techniques. *R* provides a rich set of built-in as well as extended functions for statistical, machine learning, and visualization tasks such as: data extraction, data cleaning, data loading, data transformation, statistical analysis, predictive modeling and data visualization. Currently, *R* is considered as the most popular open source and cross platform software for statistical analysis which has a very wide community support. Recently, Deep Learning [69] has been emerging a new technique which is about learning multiple levels of representation and abstraction that help to make sense of various types of data such as text, images and sound. Fig. 5 illustrates an overview of the partial landscape of big data systems. For a comprehensive survey on these systems, we refer the reader to [75].

## 6. SmartHealth framework

In general, due to recent advancements in sensor devices and other related technologies, the cost of data acquisition has reduced dramatically. In principle, while the initial setup costs are relatively high, the continuous data acquisition remains very cheap. In addition, these initial costs are continuously going down with the continuous advancements in sensor technologies. In practice, sensing-based patient monitoring generates much more data than healthcare professionals are able to manually interpret. Hence, automated processing of big data is crucially required for the sake of detecting any events of interest.

In general, electronic medical record systems (EMR) are not designed to process and handle data of large volumes, velocity and of many variety. EMR systems are also not designed to handle complex analytic operations such as anomaly detection, finding patterns in data, machine learning in addition to building complex algorithms for predictive modeling. Therefore, data analytics systems are finding their own way in the health informatics domain not just for improving patient care outcomes but also to improve the quality of care, reduce costs and improve patient population health. In particular, healthcare analytics provide methods and processes for extracting and transforming raw medical data into meaningful insight, new discovery and knowledge that supports efficient and effective healthcare decisions. Specifically, healthcare analytics goes beyond the linear and descriptive analytics which are mostly driven by the need for reporting quality measures to cover broader and deeper methods to study and analyze the data such as machine learning, non-linear algorithms in addition to the introduction of multi-analysis approaches. In general, analytics techniques can be generally categorized into the following categories [52]:

- *Descriptive analysis* is used to explain what is happening in a given situation. These techniques are commonly used to answer questions of the form "*What has happened?*" Common techniques used for this include descriptive statistics with histograms, charts, box and whisker plots, or data clustering.
- *Diagnostic analysis* is used to understand why certain things happened and what are the key drivers. For example, why is disease infection increasing? Or why are some patients readmitting every month? Common techniques for diagnostic analysis are clustering, classification, decision trees, or content analysis.
- *Predictive analysis* is used to predict what will happen in the future. It is also used to predict the probability of an uncertain outcome. For example, it can be used to predict whether a certain patient can get some diseases or not. Statistics and machine learning offer great techniques for prediction.
- *Prescriptive analysis* is used to suggest the best course of action to take to optimize your decision outcomes. Typically, prescriptive analysis combines a predictive model with business rules (e.g. decline a medication if the probability of side effect is above a given threshold). Techniques such as decision trees, linear and non-linear programming, and Monte Carlo simulation are effective here.

In order to effectively support the above analytics forms, we introduce *SmartHealth*, a framework for analytical processing of storing and processing big healthcare-related. One of the key design goals of SmartHealth is to integrate various nearby big data sources of patient data (e.g. hospital information systems (HIS), radiology information systems (RIS), laboratory information systems (LIS), picture archiving and communication systems (PACS)) in an adequate environment for applying various powerful analytics functions. Fig. 6 illustrates the layered architecture of the SmartHealth framework which capitalizes on the recent advancements on various technologies such as smart sensors, cloud computing and big data processing technologies to build a novel scalable data management and analytics platform that supports the various categories of analytics function for healthcare data sources. As shown in Fig. 6, the SmartHealth framework consists of the following main layers (from bottom to up):

- *Data Connection Layer*: This layer sets up the data sensing, collection, ingestion and pipelining steps to the centralized cloud-based data storage. The main challenge in exploiting Big Data is focused around identifying the ways of handling the diversity, heterogeneity and complexity of the data, where traditional mechanisms used for smaller datasets, e.g., manual integration or manual curation of data, are not applicable anymore, due to the volume and velocity of Big Data. This syntactic and semantic incompatibility usually results in data redundancy and inconsistencies that significantly affect the quality of the sensed data, and subsequently, the quality of the decisions taken based on this data. Semantic Web technologies (e.g. Ontologies) are the means to deal with these issues. The integration of heterogeneous multimodal medical data, including: various heterogeneous sensors, contextual information, patient's feedback, health documents, Electronic Health Records (EHR), and metadata about all connected data sources (e.g., sensors, data streams and other repositories), by means of Semantic Web technologies is one of the major problems which needs to be tackled in this layer. This includes the development and integration of relevant semantic ontologies which form the backbone of data representation and annotation. The ultimate goal of this layer is to achieve a plug-and-play compatibility of heterogeneous data sources.
- *Data Storage and Management Layer*: This layer provides a scalable, available, reliable and widely accessible data storage medium which is capable of handling massive amounts of healthcare data. This layer can be implemented using a mix of various data storage systems. For example, it can utilize scalable cloud-based relational database services (e.g. *Amazon RDS*,[39] *SQL Azure*[40]) for storing structured healthcare data while it will rely on cloud-based NoSQL storage services (e.g., *Amazon DynamoDB*,[41] *Google Datastore*[42]) for storing and processing semi-structured and unstructured data sources.
- *Analytics Layer*: This layer will provide a number of engines to provide the analytical functions. Depending on the task requirements, this layer can use one or multiple engines to execute the analytics jobs. For example, a machine learning engines (e.g., *Apache Mahout*, *SystemML*,[43] *BigML*[44]) will be concerned with the process of constructing and building adaptive models and algorithms that learn from data as well as adapt their performance as data changes over time when applied to one population for another (e.g., model that can automatically classify patients into groups that have a disease or those that do not have any disease). A predictive modeling engine will support various statistical and mathematical models to make predictions that are based on historical data. The pattern matching engine will provide tools to identify shapes and patterns in data, perform correlation analysis and clustering of data in multiple dimensions. This layer can make use of various big data analytics systems such as: Hadoop stack, Spark stack, Mahout and R.
- *Presentation Layer*: This layer will uses tools (e.g., *Tableau*,[45] *Infogram*,[46] *Plotly*[47]) for building user-friendly dashboards and applications that display the results of the analytics engine. The supported dashboards need to support various visualization schemes and be able to dynamically displaying and updating the results of the analytics jobs. Furthermore, these results drives the patient treatment process (empowered by business process management technology), which gives physicians the crucial profound insight, thus enabling them to make timely informed decisions.

In practice, the design of the SmartHealth framework is aimed at supporting the process of modifying and improving the healthcare delivery model in various ways such as:

- Relying on sensing-based screening and assessment technology in home and community environments in order to reduce the physical pressure on the environment of hospitals and turn it into an electronic flow of information.
- Changing the medication process from a reactive model to a proactive and preventative model that can significantly minimize the expenses of hospital admissions for acute events.
- Improving the personalization of the healthcare process so that individuals can monitor and identify their risk factor identification, preventative intervention, and treatment, and enables patients to live independently, while being taken care of, that has a significant positive impact on their psychological state, and subsequently on their physiological state.
- Enabling better management of clinical workloads and allowing the healthcare system to effectively prioritize the patients with the highest need.
- Supporting self-care diagnostic processes to monitor vital signs and other various measurements where these data are shared with a physician, in person or using teleconsultation, to perform a diagnosis. Furthermore, diagnosis can sometime be automated for simple illnesses such as flu.
- Optimizing the point-of-care tests by reducing the diagnosis time through minimizing the requirement to send samples away to be tested. For example, automatic testing using blood pressure cuffs and digital thermometers can support the physician to review a patient history while a measurement is being recorded.

## 7. Use cases and application scenarios

In general, the volume of healthcare data is expected to continue growing dramatically in the years ahead. In practice, utilizing the recent advancements on ICT to effectively analyze and utilize such big data can bring about significant benefits for healthcare organizations ranging from single-physician offices and multi-provider groups to large hospital networks in several use cases and application scenarios. In particular, healthcare analytics can be leveraged in several applications with the aim of turning large amounts of data into actionable information that can be exploited to identify needs, provide services, predict problems and prevent crises for the population of patients. Examples of these use cases and scenarios include [37,47,54]:

- *Big Data-Driven Decision Making*: The decision making process is based on the analysis of significant sizes of data which is more representative of the real world rather than be based

39 https://aws.amazon.com/rds/.
40 https://azure.microsoft.com/en-us/services/sql-database/.
41 https://aws.amazon.com/dynamodb/.
42 https://cloud.google.com/datastore/.
43 http://researcher.watson.ibm.com/researcher/view_group.php?id=3174.
44 https://bigml.com/.
45 www.tableau.com/.
46 https://infogr.am/.
47 https://plot.ly/.

purely on intuition. For example, the US Healthcare Big Data project [9] comprises records of over 50 million patients. This data is exploited to discover the challenges in the healthcare sector where answering the right queries for this purpose over such big data is very complex. In addition to clinical data, healthcare data also includes pharmaceutical data (e.g., drug molecules and structures), data on personal practices (e.g., exercise patterns, dietary habits, environmental factors) and billing/financial records. Effectively integrating all of this data represent a main key to significant improvements in interventions, delivery and well-being.

– *Patient profile analytics*: This involves the process of applying advanced analytics to patient profiles in order to identify patients who would benefit from proactive care or lifestyle changes. For example, applying predictive modeling techniques to model and identify the profiles of patients who are at risk of developing a specific disease (e.g., diabetes) and who should be subjected for preventive care.
– *Effective public health strategies*: This scenario includes applying analytics techniques on disease patterns in order to identify disease outbreaks and transmission for the sake of improving the performance of public health surveillance systems and their speed of response. It can be also utilized for developing faster and more effective targeted vaccination plans. In addition, big data analytics system can be utilized for capturing and analyzing social media data to predict disease outbreaks based on consumers' search, social content and query activity. For example, many researchers are currently using the *Google Trends* service[48] to study the timing and location of search engine queries to predict disease outbreaks [10,14].
– *Population management*: This scenario involves identifying potential causes for infections and readmissions. For example, identifying the most at risk patients and allocate resources wisely to help these patients (e.g., 1% of 100,000 patients had 30% of the costs). Even with structured data, standard centralized data mining can be inefficient and time consuming. Thus, distributed processing is necessary for scaling and speeding up the data mining process on such scenarios.
– *Cost Reduction*: According to US reports, approximately 5% of patients account for about 50% of all health care spending.[49] Therefore, effective prediction techniques for identifying such patients and managing them more effectively can lead to effective cost reductions. Similarly, several studies reported about the frequency and high cost of hospital readmissions [25]. Therefore, helping health care organizations to effectively predict which patients are likely to be readmitted to the hospital can provide an effective solution to reduce this cost.
– *Social network for patient*: *PatientsLikeMe*[50] is an example of a patient social network which provides an online data sharing platform that was started in 2006; now it has more than 200,000 patients and is tracking 1500 diseases. In such a platform, people connect with others who have the same disease or condition, track and share their own experiences, see what treatments have helped other patients like themselves, gain insights and identify any common patterns. In addition, patient provides the data on their individual conditions, treatment history, side effects, hospitalizations, symptoms, disease-specific functional scores, weight, mood, quality of life and more on an ongoing basis. As the demand on accessing and processing health related information from social networks is expected to increase, big data analytic systems can significantly play a major role on digesting and analyzing such increasing datasets.
– *Scalable epidemiological studies*: This types of studies enable clinicians and epidemiologists to perform large scale analysis across patient populations and care venues to help identify disease trends.
– *Evidence-based medicine*: Physicians have traditionally used their judgment when making treatment decisions, however, recently, there has been a new trend towards using evidence-based medicine. This new trend aims to optimize the decision-making process by emphasizing the use of evidence from well designed and conducted research. This trend can be supported and achieved by combining and analyzing a variety of structured and unstructured health related data, financial data, and genomic data to match treatments with outcomes, predict patients who are at risk of getting a disease and provide more efficient care.
– *Genomic analytics*: Recently, the efficiency of the process of executing gene sequencing has been significantly improved and the cost has decreased considerably. For example, the *1000 Genomes Project*[51] has been introduced as an international research effort which is coordinated by a consortium of 75 companies and organizations to establish the most detailed catalogue of human genetic variation. The project has grown to 200 terabytes of genomic data for more than 1700 persons that researchers can now freely access and analyze on Amazon Web Services for use in disease research. Thus, efficient genomic analytics techniques can make the genomic analysis process to be a main component of the regular medical care decision process and the growing patient medical record [38]. In particular, genomic analytics can play a significant role on identifying the relationships between a disease and its genetic, environmental and/or health-based risk factors. This analysis can give a unique view into the underlying mechanisms of diseases and disorders in addition to revealing the interplay between different types of risk factors. Identifying risk-based genes is a crucial mean to discover biological pathways for direct therapeutic interventions, while personal risk factors establish corrective interventions, which patients can implement to reduce their risk of developing particular diseases. Hence, leveraging effective personalized care service based on the DNA sequence information in real time can be achieved in order to highlight best practice treatments for patients. This enables moving away from a population-level epidemiological approach to small groups or individuals that are defined by their biochemistry and genetics. In addition, this contribute to achieving the ongoing and gradual shift from disease-centered care to patient-centered care. In this scenario, a single patient can have gigabytes of data in various forms including genomic, proteinic and metabolic data which is large and complex. NoSQL technology is adequate to play an effective role on dealing with such complex and massive datasets.
– *Improved remote patient monitoring*: Sensing technologies are playing a major role in improving the process of capturing and analyzing real-time and fast-moving patient data from in-hospital and in-home devices. Real time analysis of such data can significantly improve patient safety monitoring and improve the accuracy of the event prediction process. In addition, understanding how to detect reproducible patterns in signals acquired from sensing device can play an effective role on suggesting a non-invasive way of learning underlying physiological processes. Big streaming processing systems can play

48 https://www.google.com/trends/.
49 http://www.nihcm.org/images/stories/NIHCM-CostBrief-Email.pdf.
50 http://www.patientslikeme.com/.
51 http://aws.amazon.com/1000genomes/.

a major role for realizing the types of applications that can be built for addressing these scenarios.
– *Unstructured data analysis*: Nowadays, according to Gartner and IBM,[52] most data (80%) resides in unstructured or semi-structured sources where a wealth of information can be gleaned. In the healthcare systems, a lot of information are collected in clinical notes which is hard to extract. In addition, medical staff and specialist need to keep abreast of medical literature. Big data processing systems can be effectively exploited to identify highly relevant data and literature from the unstructured text and transform them into easily processable data.
– *Supporting clinical decision systems*: Clinical decision support systems have been gaining widespread attention as medical institutions and governing bodies turn towards better management of their data for effective and efficient healthcare delivery and quality assured outcomes. Analytics techniques can be exploited for processing large amounts of data, understanding, categorizing and learning from this data for the sake of predicting outcomes or recommending alternative treatments to clinicians and patients at the point of care. In addition, comparative effectiveness research can be implemented to determine more clinically relevant and cost-effective ways to diagnose and treat patients.

## 8. Conclusion and future work

In this article, we analyzed how the recent advancements of ICT can be effectively exploited and integrated for tackling the above mentioned challenges and contribute towards the state-of-the-art of healthcare services. In particular, we focused on exploiting the advancements in the areas of sensor technologies, cloud computing, Internet-Of-Things and Big data analytics systems as emerging technologies that can significantly contribute towards improving the efficiency and effectiveness of healthcare services. In addition, we proposed an integrated and comprehensive framework for big data analytics services in smart healthcare networks, *SmartHealth*, which addresses the revealed challenges and fills in the identified gaps. The framework also acts as a roadmap for future research efforts in the area of big data analytics in smart healthcare applications. Several use cases and application scenarios have been discussed to promote the importance of our proposed framework. Future work will concentrate on designing the adequate programming abstraction that can equip the analytics process for various healthcare related complex data sources (e.g., images, streams).

## References


[1] Gartner top ten disruptive technologies for 2008 to 2012: emerging trends and technologies roadshow, Technical report, Gartner, 2008.
[2] Sarfraz Alam, Mohammad M.R. Chowdhury, Josef Noll Senaas, An event-driven sensor virtualization approach for internet of things cloud, in: Proceedings of the IEEE International Conference on Networked Embedded Systems for Enterprise Applications, NESEA, 2010, pp. 1–6.
[3] Atif Alamri, Wasai Shadab Ansari, Mohammad Mehedi Hassan, M. Shamim Hossain, Abdulhameed Alelaiwi, M. Anwar Hossain, A survey on sensor-cloud: architecture, applications, and approaches, Int. J. Distrib. Sens. Netw. 2013 (2013).
[4] Arash Arami, Axelle Vallet, Kamiar Aminian, Accurate measurement of concurrent flexion–extension and internal–external rotations in smart knee prostheses, IEEE Trans. Biomed. Eng. 60 (9) (2013) 2504–2510.
[5] Michael Armbrust, O. Fox, Rean Griffith, Anthony D. Joseph, Y. Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, et al., Above the clouds: a Berkeley view of cloud computing, 2009.
[6] Luigi Atzori, Antonio Iera, Giacomo Morabito, The internet of things: a survey, Comput. Netw. 54 (15) (2010) 2787–2805.
[7] G.S. Avrunin, L.A. Clarke, L.J. Osterweil, J.M. Goldman, T. Rausch, Smart checklists for human-intensive medical systems, in: IEEE/IFIP 42nd International Conference on Dependable Systems and Networks Workshops, DSN-W, 2012, pp. 1–6.
[8] Ayan Banerjee, Sandeep K.S. Gupta, Georgios Fainekos, Georgios Varsamopoulos. Towards modeling and analysis of cyber-physical medical systems, 2011.
[9] Christian Bizer, Peter Boncz, Michael L. Brodie, Orri Erling, The meaningful use of big data: four perspectives–four challenges, ACM SIGMOD Rec. 40 (4) (2012) 56–60.
[10] John S. Brownstein, Clark C. Freifeld, Lawrence C. Madoff, Digital disease detection, harnessing the web for public health surveillance, N. Engl. J. Med. 360 (21) (2009) 2153–2157.
[11] Jason Burke, Health Analytics: Gaining the Insights to Transform Health Care, 1st edition, Wiley, July 2013.
[12] Benjamin Caballero, Theresa Clay, Sally M. Davis, Becky Ethelbah, Bonnie Holy Rock, Timothy Lohman, James Norman, Mary Story, Elaine J. Stone, Larry Stephenson, et al., Pathways: a school-based, randomized controlled trial for the prevention of obesity in American Indian schoolchildren, Am. J. Clin. Nutr. 78 (5) (2003) 1030–1038.
[13] Michele L. Callisaya, Leigh Blizzard, Michael D. Schmidt, Jennifer L. McGinley, Velandai K. Srikanth, Ageing and gait variability, a population-based study of older people, Age Ageing 39 (2010) 191–197.
[14] Herman Anthony Carneiro, Eleftherios Mylonakis, Google trends: a web-based tool for real-time surveillance of disease outbreaks, Clin. Infect. Dis. 49 (10) (2009) 1557–1564.
[15] Jeffrey Dean, Sanjay Ghemawat, Mapreduce: simplified data processing on large clusters, Commun. ACM 51 (1) (2008) 107–113.
[16] S. Don, M. Dugki, Medical cyber physical systems and bigdata platforms, in: The Medical Cyber Physical Systems Workshop, 2013.
[17] Jean P. Drouin, Viktor Hediger, Nicolaus Henke, Health care costs: a market-based view, McKinsey Q. (September 2008).
[18] Thomas Erl, Ricardo Puttini, Zaigham Mahmood, Cloud Computing: Concepts, Technology, & Architecture, Pearson Education, 2013.
[19] Hans W. Gellersen, Albercht Schmidt, Michael Beigl, Multi-sensor context-awareness in mobile devices and smart artifacts, Mob. Netw. Appl. 7 (5) (2002) 341–351.
[20] M. Goodchild, Citizens as sensors: the world of volunteered geography, GeoJournal 69 (4) (2007) 211–221.
[21] Christian Hagist, Laurence J. Kotlikoff, Health care spending: what the future will look like, 2006.
[22] Shah Ahsanul Haque, Syed Mahfuzul Aziz, Mustafizur Rahman, Review of cyber-physical system in healthcare, Int. J. Distrib. Sens. Netw. 2014 (2014).
[23] I-Min Lee, Eric J. Shiroma, Felipe Lobelo, Pekka Puska, Steven N. Blair, Peter T. Katzmarzyk, Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy, Lancet 380 (9838) (July 2012) 219–229.
[24] Lee Insup, O. Sokolsky, Chen Sanjian, J. Hatcliff, Jee Eunkyoung, Kim BaekGyu, A. King, M. Mullen-Fortino, Park Soojin, A. Roederer, K.K. Venkatasubramanian, Challenges and research directions in medical cyber physical systems, Proc. IEEE 100 (1) (2012) 75–90.
[25] Stephen F. Jencks, Mark V. Williams, Eric A. Coleman, Rehospitalizations among patients in the medicare fee-for-service program, N. Engl. J. Med. 360 (14) (2009) 1418–1428.
[26] H. Kailanto, E. Hyvarinen, J. Hyttinen, Mobile ECG measurement and analysis system using mobile phone as the base station, in: Second International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth 2008, 2008, pp. 12–14.
[27] W. Kang, Adaptive real-time data management for cyber-physical systems, Thesis, 2009.
[28] Basel Kayyali, David Knott, Steve Van Kuiken, The Big-Data Revolution in us Health Care: Accelerating Value and Innovation, Mc Kinsey & Company, 2013.
[29] D. Konstantas, R. Herzog, Continuous monitoring of vital constants for mobile users: the MobiHealth approach, in: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 4, 2003, pp. 3728–3731.
[30] Hermann Kopetz, Internet of things, in: Real-Time Systems, 2011, pp. 307–323.
[31] Tim Kraska, Ameet Talwalkar, John C. Duchi, Rean Griffith, Michael J. Franklin, Michael I. Jordan, MLbase: a distributed machine-learning system, in: CIDR, 2013.
[32] Kang Kyoung-Don, S.H. Son, J.A. Stankovic, Managing deadline miss ratio and sensor data freshness in real-time databases, IEEE Trans. Knowl. Data Eng. 16 (10) (2004) 1200–1216.
[33] Edward A. Lee, Cyber physical systems: design challenges, in: Proceedings of the 2008 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing, ISORC '08, IEEE Computer Society, Washington, DC, USA, 2008, pp. 363–369.
[34] Insup Lee, Oleg Sokolsky, Sanjian Chen, John Hatcliff, Eunkyoung Jee, Baek-Gyu Kim, Andrew King, Margaret Mullen-Fortino, Soojin Park, Alex Roederer, et al., Challenges and research directions in medical cyber-physical systems, Proc. IEEE 100 (1) (2012) 75–90.


[52] http://www.ibm.com/smarterplanet/us/en/business_analytics/article/it_business_intelligence.html.

[35] Hu Liang, Xie Nannan, Kuang Zhejun, Zhao Kuo, Review of cyber-physical system architecture, in: Proceedings of the 15th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops, ISORCW, 2012, pp. 25–30.
[36] Grzegorz Malewicz, Matthew H. Austern, Aart J.C. Bik, James C. Dehnert, Ilan Horn, Naty Leiser, Grzegorz Czajkowski, Pregel: a system for large-scale graph processing, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 2010, pp. 135–146.
[37] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, McKinsey Global Institute, Big data: the next frontier for innovation, competition, and productivity, 2011.
[38] Vivien Marx, Biology: the big challenges of big data, Nature 498 (7453) (2013) 255–260.
[39] Fabien Massé, Martien Van Bussel, Aline Serteyn, Johan Arends, Julien Penders, Miniaturized wireless ECG monitor for real-time detection of epileptic seizures, ACM Trans. Embed. Comput. Syst. 12 (4) (2013) 102.
[40] Colm McCaffrey, Olivier Chevalerias, Cian O'Mathuna, Karen Twomey, Swallowable-capsule technology, IEEE Pervasive Comput. 7 (1) (2008) 23–29.
[41] Michael J. McGrath, Cliodhna Ni Scanaill, Sensor Technologies: Healthcare, Wellness and Environmental Applications, Apress, 2013.
[42] Michael J. McLachlan, John A. Katzenellenbogen, Huimin Zhao, A new fluorescence complementation biosensor for detection of estrogenic compounds, Biotechnol. Bioeng. 108 (12) (2011) 2794–2803.
[43] Sergey Melnik, Andrey Gubarev, Jing Jing Long, Geoffrey Romer, Shiva Shivakumar, Matt Tolton, Theo Vassilakis, Dremel: interactive analysis of web-scale datasets, Proc. VLDB Endow. 3 (1–2) (2010) 330–339.
[44] E.O. Mendez, Ren Shangping, Design of cyber-physical interface for automated vital signs reading in electronic medical records systems, in: IEEE International Conference on Electro/Information Technology, EIT, 2012, pp. 1–10.
[45] Jing Meng, Dong Liang, Liang Song, Compressed sensing photoacoustic tomography in vivo in time and frequency domains, in: IEEE–EMBS International Conference on Biomedical and Health Informatics, BHI, 2012, pp. 717–720.
[46] Robert E. Miller, Keith S. Reisinger, Mark M. Blatter, Frederick Wucher, Pediatric counseling and subsequent use of smoke detectors, Am. J. Publ. Health 72 (4) (1982) 392–393.
[47] Travis B. Murdoch, Allan S. Detsky, The inevitable application of big data to health care, JAMA 309 (13) (2013) 1351–1352.
[48] Huansheng Ning, Unit and Ubiquitous Internet of Things, CRC Press, 2013.
[49] Sungmee Park, Sundaresan Jayaraman, Enhancing the quality of life through wearable technology, IEEE Eng. Med. Biol. Mag. 22 (3) (2003) 41–48.
[50] Andrew Pavlo, Erik Paulson, Alexander Rasin, Daniel J. Abadi, David J. DeWitt, Samuel Madden, Michael Stonebraker, A comparison of approaches to large-scale data analysis, in: Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, 2009, pp. 165–178.
[51] Daryl C. Plummer, Thomas J. Bittman, Tom Austin, David W. Cearley, David Mitchell Smith, Cloud computing: defining and describing an emerging phenomenon, Gartner 17 (2008).
[52] Foster Provost, Tom Fawcett, Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking, O'Reilly Media, 2013.
[53] M. Quin, Predicting vital signs, 2014.
[54] Wullianallur Raghupathi, Viju Raghupathi, Big data analytics in healthcare: promise and potential, Health Inf. Sci. Syst. 2 (1) (2014) 3.
[55] Sherif Sakr, Mohamed Gaber, Large Scale and Big Data: Processing and Management, CRC Press, 2014.
[56] Sherif Sakr, Anna Liu, Daniel M. Batista, Mohammad Alomari, A survey of large scale data management approaches in cloud environments, IEEE Commun. Surv. Tutor. 13 (3) (2011) 311–336.
[57] Sherif Sakr, Anna Liu, Ayman G. Fayoumi, The family of mapreduce and large-scale data processing systems, ACM Comput. Surv. 46 (1) (2013) 11.
[58] Amit Sheth, Pramod Anantharam, Cory Henson, Physical-cyber-social computing: an early 21st century approach, IEEE Intell. Syst. 28 (2013) 78–82.
[59] Victor Shnayder, Bor-rong Chen, Konrad Lorincz, Thaddeus R.F. Fulford Jones, Matt Welsh, Sensor networks for medical care, 2005.
[60] James Staten, Simon Yates, Frank E. Gillett, Walid Saleh, Rachel A. Dines, Is cloud computing ready for the enterprise, Forrester Res. 7 (2008).
[61] Sameer Tilak, Real-world deployments of participatory sensing applications: current trends and future directions, ISRN Sens. Netw. 2013 (2013).
[62] Stephen Uzor, Lynne Baillie, Exploring & designing tools to enhance falls rehabilitation in the home, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013, pp. 1233–1242.
[63] Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, Maik Lindner, A break in the clouds: towards a cloud definition, Comput. Commun. Rev. 39 (1) (2008) 50–55.
[64] Paul Vegoda, Introducing the IHE (integrating the healthcare enterprise) concept, J. Healthc. Inf. Manag. 16 (1) (2002) 22.
[65] K.K. Venkatasubramanian, S. Nabar, S.K.S. Gupta, R. Poovendran, Cyber physical security solutions for pervasive health monitoring systems, in: E-Healthcare Systems and Wireless Communications: Current and Future Challenges, IGI Global, 2012.
[66] J. Wang, H. Abid, S. Lee, L. Shu, F. Xia, A secured health care application architecture for cyber-physical systems, Control Eng. Appl. Inform. 13 (3) (2011) 101–108.
[67] Jin Wang, Hassan Abid, Sungyoung Lee, Lei Shu, Feng Xia, A secured health care application architecture for cyber-physical systems, arXiv:1201.0213, 2011.
[68] Rolf H. Weber, Romana Weber, Internet of Things, Springer, 2010.
[69] Van B. Weigel, Deep Learning for a Digital Age: Technology's Untapped Potential to Enrich Higher Education, ERIC, 2002.
[70] A. Wood, J.A. Stankovic, G. Virone, L. Selavo, He Zhimin, Qiuhua Cao, Thao Doan, Wu Yafeng, Fang Lei, R. Stoleru, Context-aware wireless sensor networks for assisted living and residential monitoring, IEEE Netw. 22 (4) (2008) 26–33.
[71] Fang-Jing Wu, Yu-Fen Kao, Yu-Chee Tseng, From wireless sensor networks towards cyber physical systems, Pervasive Mob. Comput. 7 (4) (2011) 397–413.
[72] Chin-Lung Yang, Song-Jeng Huang, Cheng-Wei Chou, Yan-Chao Chiou, Kun-Piao Lin, Min-Shiun Tsai, Kung-Chia Young, Design and evaluation of a portable optical-based biosensor for testing whole blood prothrombin time, Talanta 116 (2013) 704–711.
[73] Li Yang, Wu Chao, Guo Li, Lee Chun-Hsiang, Guo Yike, Wiki-Health: A Big Data Platform for Health Sensor Data Management, IGI Global, Hershey, PA, USA, 2014, pp. 59–77.
[74] Madoka Yuriyama, Takayuki Kushida, Sensor-cloud infrastructure – physical sensor management with virtualized sensors on cloud computing, in: Proceedings of the 13th International Conference on Network-Based Information Systems, NBiS, 2010, pp. 1–8.
[75] Hao Zhang, Gang Chen, Beng Chin Ooi, Kian-Lee Tan, Meihui Zhang, In-memory big data management and processing: a survey, IEEE Trans. Knowl. Data Eng. 27 (7) (2015) 1920–1948.
[76] Zhang Zhaoyang, Wang Honggang, Wang Chonggang, Fang Hua, Interference mitigation for cyber-physical wireless body area network system using social networks, IEEE Trans. Emerg. Top. Comput. 1 (1) (2013) 121–132.