# Big Data HIS of the IRCCS-ME Future: The Osmotic Computing Infrastructure

Lorenzo Carnevale[1,2](✉), Antonino Galletta[1,2], Antonio Celesti[1], Maria Fazio[1], Maurizio Paone[2], Placido Bramanti[2], and Massimo Villari[1,2]

[1] Department of Engineering, University of Messina, Messina, Italy
{lcarnevale,angalletta,acelesti,mfazio,mvillari}@unime.it
[2] IRCCS Centro Neurolesi "Bonino Pulejo", Messina, Italy
{lcarnevale,angalletta,mpaone,pbramanti,mvillari}@irccsme.it

**Abstract.** Nowadays, we are observing a massive digitalization of clinical tasks in Hospital Information Systems (HIS). Even more medical devices belongs to Internet of Things (IoT) applications that generate a huge amount of clinical data. Therefore, the healthcare industry is looking at modern big data storage, processing and analytics technologies. In this context, traditional HIS presents several issues, such as a mismanagement of data generated from medical devices. Starting from the experience of IRCCS Centro Neurolesi "Bonino Pulejo" placed in Messina (Italy), i.e., a clinical and research center, in this paper, we motivate the need to move traditional HIS into an innovative infrastructure based on the Osmotic computing paradigm. In particular, studying the healthcare domain, we specifically focus on production and research tasks.

**Keywords:** Hospital Information System · Big Data · eHealth · IoT · Osmotic computing

## 1 Introduction

Nowadays, the healthcare industry is facing many challenges such as waste reduction, integration of a new generation of electronical medical systems, collection and communication of a huge amount of clinical data in a quick and safe fashion. Up to now, for the healthcare industry has not been easy to introduce new technological improvements in the daily work of the clinical personnel but, currently, medical and governmental authorities of many countries encourage the adoption of cutting-edge information technology solutions in healthcare. Typical examples of famous initiatives include Electronic Health Record (EHR), Remote Patient Monitoring (RPM) and tools for medical decision making.

The healthcare industry is looking at modern Big Data storage, processing, and analytics technologies. An analysis of the McKinsey Global Institute [1] studied the Big Data penetration for healthcare, highlighting a good potential to achieve insights. In the period between 2010 and 2015, it measured that the


A. Longo et al. (Eds.): IISSC 2017/CN4IoT 2017, LNICST 189, pp. 199–207, 2018.
https://doi.org/10.1007/978-3-319-67636-4_21

number of connected nodes increased up to 50 million. In the same period, the Organisation for Economic Co-operation and Development (OECD) [2] measured about $100 billion, as a share of Gross Domestic Product (GDP) and total health spending, for the final consumption of healthcare goods and services (investments not included). The estimate takes into account both private and public financing. Moreover, in [3], Forbes highlights how Big Data is changing healthcare. Indeed, in addition to economical healthcare perspective, such as improving profits and cutting down on wasted overhead, healthcare operators are looking at Big Data analytics in the context of epidemics prediction and disease care in order to improve the quality of life of patients and avoid preventable deaths. New models of treatment delivery are rapidly changing because population and longevity grow and the decisions making about these changes are driven by data. Nowadays, the trend is to understand as much as possible the patients' status, collecting signals about serious illness in a preliminary stage in order to ensure simpler and cheaper treatment.

In this context, many healthcare organizations have evaluated the possible adoption of Cloud Computing solutions due to the intrinsic ability to access services through standard mechanisms that promote heterogeneous clients utilization and elastically provide services able to spatially and temporally scale up/down without limits [4]. Medical devices have been seen as part of Internet of Things (IoT) [5] applications either directly connected to the network or indirectly connected to the Hospital Information System (HIS) through clinical workflows. Nevertheless, this is still not enough. The emerging availability and complexity of various types of medical devices, along with large data volumes that such devices generate, can limit the current Cloud-centric IoT programming models. Thus, the current systems need to be revised into something that is more adaptable and decentralized in order to meet the emerging healthcare applications needs. The Osmotic Computing [6] paradigms aims to decompose applications into microservices [7,8] and to perform dynamic tailoring of microservices in smart environments exploiting resources in Edge and Cloud infrastructures. Application delivery follows an Osmotic behavior where microservices in containers are deployed opportunistically in Cloud and Edge systems.

This scientific work highlights the experience reported at IRCCS Centro Neurolesi "Bonino Pulejo" (Messina, Italy), a healthcare clinical and research centre. In particular, we talk about motivations that led the IRCCS' managers to invest in the modernization of their IT infrastructure. This experience has shown the importance of a lean planning in order to meet demanding requirements both for productive and research areas of dynamic healthcare environments. Moreover, we highlight the Osmotic Computing advantages for the creation of Big healthcare Data infrastructure.

The rest of the paper is organized as follows. Related works are summarized in Sect. 2. Section 3 discusses the reasons because a HIS should adopt Cloud technologies. The architecture is presented in the Sect. 4, whereas the aspects related to Osmotic Computing applied to Big healthcare Data analytics are presented in Sect. 5. Conclusion and lights to the future are presented in Sect. 6.

## 2 Related Work

Nowadays, Cloud computing into the healthcare domain is a really challenging topic. To demonstrate this, in the following we report several scientific works that aim to improve HIS solutions. In [9] authors focused on Digital Imaging and Communications in Medicine (DICOM), a standard for storing and managing medical images. They proposed a hybrid model for Cloud-based HIS focusing about public and private Cloud. The last one was able to manage and store computer-based patients records and other important information, while a public Cloud was able to handle management and business data. In order to share information among different hospitals, authors adopted VPN (Virtual Private Network). In [10] authors proposed a Cloud-based Tele-Medicine system that, thanks to wearable sensors, allows patients to send eHealth data, such as Blood Pressure (BP) and Electrocardiogram (ECG), into specific gateways in order to forward them to the Cloud. Here, data were processed and compared with existing results. If system founds suitable results, it sends back an automatic feedback, otherwise the appropriate physician is intimated via phone call or SMS (Short Message System). Using their PDA/smartphone, physicians can get patients' data in order to diagnose the disease and send back reports. Four services for a Healthcare as a Service (HaaS) model have been proposed in [11]:

1. Storage Archival and Indexing Services: using features of commercial Cloud providers (like Amazon, Microsoft Azure, etc.), authors proposed a Cloud-based Picture Archiving and Communication System (PACS) to stored both Medical Images and patient's personal data;
2. Image Processing Services: authors proposed an image processing service on a Virtual Machine that is able to retrieve images from Cloud storage, compute and encapsulate them into JavaScript Object Notation (JSON) format and, moreover, to send results to client applications.
3. Reporting Services: a mechanism to share medical reports among different hospitals.
4. Charting and Trend Analysis of Healthcare Data: a system that aggregates medical data and performs different kinds of analysis.

In [12] the authors presented a hybrid storage solution for the management of eHealth data, which proved the synergetic utilization of SQL-like strategies ad NoSQL document based approaches. Specifically, this scientific work adopted the proposed solution for neurologic Tele-Rehabilitation (TR) of patients at home. A Open Archivial Storage System (OAIS) for Cloud-based HIS able to manage Big Clinical Data through a NoSQL column oriented approach was presented in [13]. Finally, a Cloud-based next-generation sequencing (NGS) tool has been proposed in [14]. The authors investigated the existing NGS solutions, highlighting the necessary missing features in order to move toward the achievement of an ecosystem of biotechnology Clouds.

## 3 Motivation

Nowadays, traditional HIS are composed of several independent subsystems that perform specific tasks and store personal and medical patients' data into local repositories. Nevertheless, this kind of configuration presents several issues for patient, physician, technician and administrative staff. From the patient point of view, the management of exams outcomes is difficult. They can not retrieve results of their analysis. Indeed, at present, they have to request these to administrative staff of each ward. The mismanagement of HIS causes losing of a huge amount of clinical data generated both by human and machine sources. In this way, only a small quantity of gathered data can be analysed. Moreover, this quantity is managed through spreadsheets, causing difficulties to correlate clinical data and find out insights. As mentioned above, HIS is composed of several black-box subsystems that require specific servers and hardware/software configurations. This kind of infrastructure is really expensive for hardware, cooling and powering costs. Furthermore, it is difficult to be managed due to the fact that update operations are hard to be accomplished because system administrators have to replicate the same tasks several times.

## 4 IRCCS' Infrastructure at a Glance

The objective of this scientific work is to describe a Cloud-based HIS, which integrates daily clinical activities along with digitalization and analysis processes in order to support healthcare professionals. Our case study is the IT infrastructure of IRCCS Centro Neurolesi "Bonino Pulejo" placed in Messina (Italy). It is a scientific institute for recovery and care with the mission in the field of neuroscience for the prevention, recovery and treatment of individuals with severe acquired brain injuries, besides spinal cord and neurodegenerative diseases, by integrating highly specialized healthcare, technological innovation and higher education. To this end, the clinical activities need to be divided into two categories: Production and Research. With reference to Fig. 1, the production side includes all services that facilitate the administrative and healthcare personnel; on the other hand, the research side includes all innovative services that support the IRCCS' healthcare research activities. Moreover, for each category, we thought several thematic areas such as Frontend & Communication, Security & Privacy, Microservices, Big Data and Storage.

All these services are supported from a powerful physical infrastructure. With reference to Fig. 2, the infrastructure is largely built using traditional systems because it is required to build a solid foundation for the production line. At the same time, research line is supported by the same infrastructure. It is composed by three layers (Storage, Network and Computation) linked together. Storage disks are network available thanks to the iSCSI internet protocol. Instead, computation is provided with different technologies. Xen Server is the hypervisor used to provide traditional virtual environment and it allows to run several Virtual Machine for different purposes. Docker and Kubernates provide a lighter

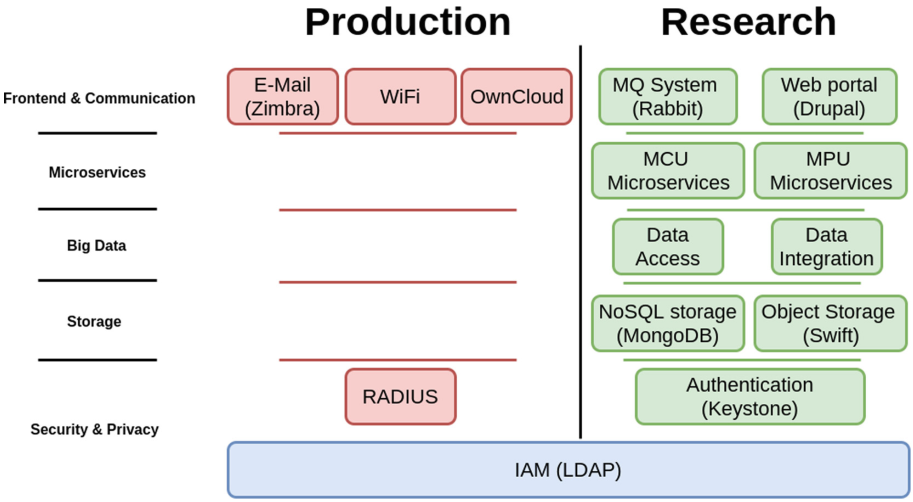


**Fig. 1.** Services list of IRCCS Institute.

virtual environment for each specific purpose. Finally, microcontroller devices mount firmware and provide another way to compute a specific purpose. All of these support RESTful/MQTT clients or servers for input/output network communications.

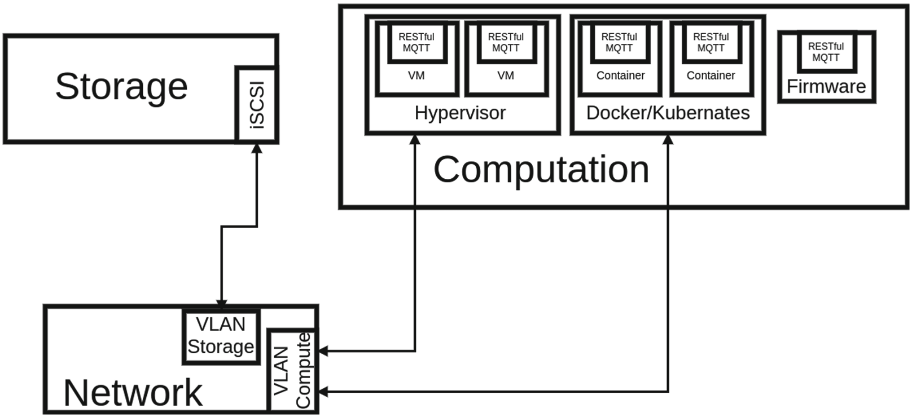


**Fig. 2.** Layers infrastructure of IRCCS institute

Following this section, we deepen the implemented services both for production and research sides. The only one service shared for both sides was the Identity and Access Management (IAM). It allows to manage users' authentication and authorization for the whole healthcare system. Indeed, thanks to

Lightweight Directory Access Protocol (LDAP), our IAM system provides unique credentials for all users both for productive and research Cloud services. In the following, we discuss in detail both production and research services.

### 4.1 Production Services

From the healthcare perspective, the main constraint for new services' production for patients care clinical support is given by the current national and international regulations. Therefore, our assistance was verifiable exclusively in the Frontend & Communication area in order to facilitate healthcare and administrative activities in low patient risk context or, if necessary, be used by patients and family members to improve quality of service during the time spent into the hospital.

As support for IAM, productive Cloud-based architecture includes RADIUS (Remote Authentication Dial-In User Service), an authentication, authorization and accounting protocol in order to give network access. Indeed, on top of this, a wireless communication was developed. The introduction of a stable and extensive wireless network, where there was only a wired network, has allowed the growth of heterogeneous devices connected to the Internet, besides to facilitate services access from any local Institute.

Other powerful tools for daily activities have been business email and web storage services. The first one is a Zimbra-based web mail software for messaging and collaboration; while the web storage services has been deployed using ownCloud, "an open source, self-hosted file sync and share app platform" [15]. In order to facilitate the utilization of both of them, we integrated them into a single Software as a Service (SaaS) web tool accessible from any browser using unique credentials stored on IAM. The advantages to the medical staff has been immediate. Indeed, thanks to specific functionalities such as calendar management and sharing files, it was possible to integrate the turn calendar within.

### 4.2 Research Services

Compared to productive side, the research side has had greater freedom of thought and development because it is not affected by regulatory constraints. The advantage is quickly evidenced by the multitude of services implemented and visible on the right side of Fig. 1. Moreover, each implemented research service follows a well-defined line: Big Data storage and management.

According to [16], the best approach to be adopted in the Big Data infrastructure design for healthcare is suggested by the lean philosophy. Indeed, after each software release, we provided research tools to healthcare personnel and gathered feedback useful to re-address our work. Therefore, we started with a well defined sandbox.

The first challenge was to acquire all data coming from medical devices considering that the low IT expertise of the clinical personnel can (potentially) be a limit. In this regard, our goal was to minimize the number of IT operations they have to perform. The first step of this clinical workflow was to move outcomes of medical devices inside shared directories. It is the only requirement requested by healthcare professionals. Data Access service continuously listen on shared directories to monitor the addition of new outcomes. Data collected from shared directories are parsed and integrated with unstructured storage systems such as MongoDB, i.e., a scalable NoSQL database. In order to accomplish this task we provided the Data Integration service. Moreover, as support of this clinical workflow we thought a message queue system such as RabbitMQ, which task is to maintain data acquisition requests in an ordered list of messages.

Along with text/numbers data, the pool of main generated data coming from healthcare industry includes also images [1], such as Magnetic Resonance Imaging (MRI), radiograph films, etc. In this domain file systems or block storage architecture have management limits because they manage, respectively, data as a file hierarchy and data as blocks with sectors and tracks. Instead, object storage manages data as object, including data itself, metadata and globally unique identifiers. For this purpose, our research Cloud-based architecture includes Swift's OpenStack as object storage, "a highly available, distributed, eventually consistent object store" [17]. This wide uses another OpenStack service for authentication and high-level authorization, namely Keystone. It supports integration with LDAP directories, extending the Identity and Access Management discussed in starting part of the Sect. 4.

Other clinical data and services are provided by MCU and MPU controllers. Their task is to extend the functionalities of medical devices in order to complete the set of data associated to patients during analysis. Thus, they are typical IoT applications for specific operations.

Finally, all services are available through a Drupal-based web portal. It allows to manage each service in terms of configuration and visualization.

## 5 Osmotic Evolution of the Infrastructure

The discussion in the previous Section, specially regarding to the set of services offered by the research branch of Cloud-based architecture, provides the foundations for a well-defined Big Data storage and management. As a result, we approve the creation of a new model service in addition to the three models (Infrastructure as a Service, Platform as a Service and Software as a Service) theorized by National Institute of Standards and Technology (NIST). This model is called Big Data as a Service and includes all operations useful to discover insights, from raw data access produced by medical devices and personnel up to displaying analysed data. More specifically, it includes several phases such as asynchronous and real-time data access, data integration and normalization, data cleaning in order to remove errors, inaccuracies or inconsistent data, data analysis and visualization of final outcomes. Moreover, we must not neglect data access security mechanisms. Among these phases, some are automated and repeatable while others have a different cycle of utilization and require more human interactions. Furthermore, in order to complete all Big Data operations, we need to add other technologies such as Hadoop and Spark.

The characteristic shared by these phases is that each of them can be handled as a stand alone system but, at the same time, as part of a well-defined healthcare workflow. Thus, it is easy to think about phases as microservices that are dynamically tailored on hosting smart environments. The Osmotic Computing was born from the dynamic movement of microservices that individually perform their tasks, but together complete an ensemble action. Indeed, like the movement of solvent molecules through a semipermeable membrane into a region of higher solute concentration to equalize the solute concentrations on the two sides of the membrane - that is, osmosis (in the context of chemistry) - in Osmotic Computing, the dynamic management of resources in Cloud and Edge datacenters evolves toward the balanced deployment of microservices satisfying well-defined low-level constrains and high-level needs. However, unlike the chemical osmotic process, Osmotic Computing allows a tunable configuration of involved resources, following resource availability and application requirements.

The advent of Osmotic Computing as management paradigm of microservices has been enabled by the proliferation of light virtualization technologies (such as Docker and Kubernates), alternatively to traditional approaches based on hypervisor (such as Xen and VMWare). Adapting microservices to the physical characteristics of underlying infrastructure by using decision making strategies that map them, the Osmotic Computing reduces waste in terms of systems administration and energy costs.

## 6 Conclusion and Future Work

In this paper, we discussed a novel technological approach to deploy a HIS able to store Big eHealth Data collected through several medical devices. In particular, the paper presents a collection of motivations in order to migrate from traditional legacy HIS to Osmotic-based HIS solutions, improving data sharing, energy management and system administration.

In order to accomplish these goals, our idea was to create a robust infrastructure based on new Osmotic Computing paradigm, implementing all hospital operations as specific microservices. Moreover, studying the healthcare domain, we decided to divide our infrastructure into two different branches: one for production and another one for research tasks. The first one offers facilities to healthcare and administrative personnel, while the latter exposes innovative tools to support mostly Big Data research activities. As union point between both sides, we deployed an IAM service in order to provide a unique authentication and authorization system.

Given the infrastructural platform built and described in this paper, in our future work we aim to design an agile Osmotic-based piece of middleware able to perform microservices for Big healthcare Data management.

**Acknowledgment.** The authors would like to thank IRCCS Centro Neurolesi "Bonino Pulejo" and all healthcare personnel that helped us during design and deploy of the infrastructure.

## References


1. Big Data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, June 2011
2. OECD, Health spending (indicator) (2016). http://dx.doi.org/10.1787/8643de7e-en
3. How big data is changing healthcare. https://goo.gl/R0RIOb
4. Celesti, A., Peditto, N., Verboso, F., Villari, M., Puliafito, A.: Draco paas: a distributed resilient adaptable cloud oriented platform. In: 2013 IEEE International Symposium on Parallel Distributed Processing, Workshops and Phd Forum, pp. 1490–1497, May 2013
5. Celesti, A., Fazio, M., Giacobbe, M., Puliafito, A., Villari, M.: Characterizing cloud federation in IoT. In: 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 93–98, March 2016
6. Villari, M., Fazio, M., Dustdar, S., Rana, O., Ranjan, R.: Osmotic computing: a new paradigm for edge/cloud integration. IEEE Cloud Comput. **3**(6), 76–83 (2016)
7. Enabling microservices: containers & orchestration explained, July 2016. https://www.mongodb.com/collateral/microservices-containers-and-orchestration-explained
8. Microservices: The evolution of building modern applications, July 2016. https://www.mongodb.com/collateral/microservices-the-evolution-of-building-modern-applications
9. He, C., Jin, X., Zhao, Z., Xiang, T.: A cloud computing solution for hospital information system. In: 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, vol. 2, pp. 517–520, October 2010
10. Parane, K.A., Patil, N.C., Poojara, S.R., Kamble, T.S.: Cloud based intelligent healthcare monitoring system. In: 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 697–701, February 2014
11. John, N., Shenoy, S.: Health cloud - healthcare as a service (HaaS). In: 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1963–1966, September 2014
12. Fazio, M., Bramanti, A., Celesti, A., Bramanti, P., Villari, M.: A hybrid storage service for the management of big e-health data: a tele-rehabilitation case of study. In: Proceedings of the 12th ACM Symposium on QoS and Security for Wireless and Mobile Networks, pp. 1–8 (2016)
13. Celesti, A., Maria, F., Romano, A., Bramanti, A., Bramanti, P., Villari, M.: An oais-based hospital information system on the cloud: analysis of a NoSQL column-oriented approach. IEEE J. Biomed. Health Inform. **99**, 1 (2017)
14. Celesti, A., Celesti, F., Fazio, M., Bramanti, P., Villari, M.: Are next-generation sequencing tools ready for the cloud? In: Trends in Biotechnology (2017)
15. ownCloud. http://www.owncloud.org
16. The Big Big Data Workbook, Informatica (2016)
17. Swifts documentation. http://www.docs.openstack.org/developer/swift