



## CLOUD COMPUTING IN HEALTHCARE AND BIOMEDICINE

BARBARA CALABRESE\* AND MARIO CANNATARO†

**Abstract.** High throughput platforms available in clinical settings or in research laboratories, such as magnetic resonance imaging, microarray, mass spectrometry and next-generation sequencing, are producing an increasing volume of clinical and omics data that poses new issues in terms of secure data storage, models for data integration and analysis, and high performance computing. Cloud Computing offers large scalable computing and storage, data sharing, on-demand anytime and anywhere access to resources and applications, and it supports easy but powerful distributed computing models, for facing those issues. In fact, in the recent years it has been adopted for the deployment of several applications in healthcare and bioinformatics both in academia and in the industry. However, cloud computing presents several issues regarding the security and privacy of data, that are particularly important when analyzing patients data, such as in personalized medicine. This paper reviews main cloud-based healthcare and biomedicine applications; with a special focus on healthcare, biomedicine and bioinformatics solutions and underlines main issues and problems related to the use of such platforms for the storage and analysis of patients data.

**Key words:** cloud computing, healthcare, biomedicine, bioinformatics, virtualization

**AMS subject classifications.** 68M14, 68U01

**1. Introduction.** Research in life sciences, as well as, clinical practice in medicine, are more and more based on the large datasets produced by high throughput platforms used for the investigation of the human body (e.g. magnetic resonance imaging), as well as the cell machinery (e.g mass spectrometry, microarray, and next generation sequencing). For instance, medical imaging is producing large datasets of biomedical images, while omics sciences, such as genomics, proteomics and interactomics, are producing large datasets of experimental data. Some recent advances in biomedicine and healthcare are based on the integrated analysis of clinical and omics data, as in pharmacogenomics. Pharmacogenomics is an important branch of genomics that studies the impact of genetic variation (e.g. Single Nucleotide Polimorphisms - SNPs) on drug response in patients and is at the basis of the so-called "personalized medicine", where drugs are chosen or optimized to meet the genetic profile of each patient. The application of omics studies on large populations is producing an increasing amount of experimental and clinical data, as well as specialized databases spread over the Internet. However, the storage, preprocessing and analysis of experimental data are becoming the main bottleneck of the analysis pipeline.

Cloud computing is a computing model that has spread very rapidly in recent years for the supply of IT resources (hardware and software) of different nature, through services accessible via the network. Cloud can provide Virtual Machines (VM) in the sense of hardware resources, platforms to deploy applications or ready-to-use services. The cloud resources are dynamically scalable, virtualized and accessible on the Internet [1]. This model provides new advantages related to massive and scalable computing resources available on demand, virtualization technology and payment for use as needed [2].

Thus, cloud computing may play an important role in many phases of the bioinformatics analysis pipeline, from data management and processing, to data integration and analysis, including data exploration and visualization.

Healthcare requires continuous and systematic innovation in order to remain cost effective, efficient and to provide high-quality services. Many managers and experts predict that cloud computing can improve healthcare services and benefit healthcare research. The work [3] summarizes main opportunities and challenges of cloud computing to improve healthcare services. The author recalls that Schweitzer [4], Haughton [5], and Kabachinski [6] believe that cloud computing can reduce electronic health record (EHR) startup expenses, and licensing fees, and thus this will encourage cloud computing adoption.

Different works reported the successful application of cloud computing in healthcare and biomedicine [7, 8,

---

\*Laboratory of Bioinformatics, Department of Medical and Surgical Sciences, University "Magna Graecia", viale Europa 88100, Catanzaro, Italy ([calabreseb@unicz.it](mailto:calabreseb@unicz.it)).

†Corresponding Author. Laboratory of Bioinformatics, Department of Medical and Surgical Sciences, University "Magna Graecia", viale Europa 88100, Catanzaro, Italy ([cannataro@unicz.it](mailto:cannataro@unicz.it)).

9, 10]. Besides academic researchers, many world-class software companies have heavily invested in the cloud, such as Microsoft HealthVault ([www.healthvault.com](http://www.healthvault.com)).

Despite the many benefits associated with cloud computing to healthcare, there are also several management, technology, security, and legal issues to be addressed. For example, the storing of personal health information into a party, remote data center raises serious issues related to patient privacy. The possibility that patient data could be lost, misused or fall into the wrong hands, could affect rapid adoption.

The aim of this paper is to describe and discuss the most significant applications of cloud computing in the healthcare and biomedicine. The examined fields include bioinformatics, medical informatics, telemedicine and bioimaging. Then the paper focuses on specific requests and issues of such applications on cloud computing. The paper is organized as follows: in the Section 2 cloud computing definition is discussed. Service and delivery models are presented in order to define the cloud-related background. Successively, in the Section 3 the paper focuses on the application of cloud computing in bioinformatics, molecular modeling, medical informatics (EHR), medical imaging, telemedicine. Section 4 summarizes the main problems to be faced when moving biomedical applications on the cloud. Finally, Section 5 concludes the paper and outlines future developments.

**2. Background on Cloud Computing.** Even though cloud computing is now becoming the key technology for the storage and analysis of large data sets both in academia and industry, it is not a totally new concept. In fact, it has some relations with grid computing and other technologies such as utility computing, clustering, virtualization systems, and distributed systems.

Specifically, High Performance Computing (HPC) covers all aspects of computing that require a lot of computing power and memory [11]. Common HPC computer environments are multi-core processors, clusters, grids, graphics processing units (GPUs). These computing environments have one thing in common: serial code will not run faster on these computers than on any other normal computer. The power of HPC environments is in fact gained through the possibility of parallel or distributed computing. In other words, exploiting HPC requires extensive rewriting of serial code. On the other hand, cloud computing can be interpreted as the next generation of grid computing because the delivery of computing is seen as a service rather than a product, and moreover, it supports easy to implement distributed computing models such as map-reduce [12].

There are many definitions of cloud computing and in particular most of the definitions and background material described in the rest of this section were taken from the work by Ebejer et al [54]. The first definition comes from the work of Mell and Grance [1] and is a popular working definition of cloud computing from the National Institute of Standards and Technology, U.S. Department of Commerce. Their definition focuses on computing resources which can be accessed from anywhere and may be provisioned online. It also specifies five characteristics of cloud computing (i.e. on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service), three service models (i.e. Software as a Service, Platform as a Service and Infrastructure as a Service) and four deployment methods (i.e. private cloud, community cloud, public cloud and hybrid cloud). Most of the other definitions do not mention deployment methods. In contrast to other definitions, this one does not explicitly mention virtualization as a key technology.

According to this definition, the five key characteristics of cloud services are:

- on demand self service: a consumer can acquire computing power unilaterally and automatically without requiring any human intervention by the service provider;
- broad network access: the computational capabilities are accessible on the Internet in accordance with standard mechanisms;
- resource pooling: users have the impression that the available resources are unlimited and can be purchased in any amount and at any time. Resources of the service provider come together to serve a variety of consumers, according to a multi-tenant model. The physical and virtual resources are dynamically assigned and reassigned to consumer, based on his/her requests without having any control or knowledge of the exact location of the resources assigned to him/her;
- rapid elasticity: resources are able to be quickly allocated and elastically;
- measured services: the cloud systems automatically control and optimize resource use by evaluating appropriate parameters (e.g, storage, processing power, bandwidth and active user accounts).

Buyya et al. [13] supply a definition with the vision that cloud computing will become the fifth utility and that cloud computing is the next generation data centre. This definition extensively highlights the importance

of Service Level Agreement (SLA) and the lack of market oriented resource management (e.g. violation of SLAs, automatic allocation of resources to fulfil SLAs, etc.) in cloud environments.

Grossman [14] offers a general, working definition of cloud computing, which highlights that cloud computing environments should behave like local data centres.

Vaquero et al. [15] collected 22 excerpts from previous works and fused these into a single definition by studying the common properties of cloud computing. This definition emphasizes the importance of Service Level Agreements (SLA) in order to increase confidence in the cloud environment and defines virtualization as the key enabler of cloud computing.

In the paper [16], Marston et al. explicitly state that cloud computing lowers costs by offsetting capital costs with operational ones. Moreover, they give an interesting business perspective to cloud computing which is missing in almost all the other related publications.

The definition taken from [17] by Intel, offers a particular perspective on cloud computing, from the vantage point of a major hardware manufacturer with a commercial interest in building clouds. It describes three main aspects that help realize cloud computing: *federation* (ease of moving data and services within the cloud), *automation* (cloud services should be provisioned with no human intervention) and *client-awareness* (cloud-based applications can take advantage of the capabilities of the end-point device). The definition also makes explicit reference to security, by evidencing that only authenticated devices may connect to the cloud environment.

Hill et al. [18] give a recent definition of cloud computing that presents cloud computing as a utility.

Finally, the Open Grid Forum (OGF) reports a cloud definition in [19]. Specifically, cloud is a data processing infrastructure providing an extensive integrated client interface permitting information discovery, data management, job management, logging, accounting and monitoring. Most clouds are enclosed inside a single administrative domain, do not provision communication between different administrative domains, and provide limited security.

Thus, it is possible to summarize the main advantages of cloud computing model:

- Service-oriented: computing services are all offered as end-products to customers. Therefore, the administration and maintenance of these devices fall under the responsibility of the cloud vendor.
- Massive scalability: the allocation of resources is flexible and elastic, allowing an efficient management of spikes in computational requirements.
- On-demand resources: resources can be added in real time, sometimes in an automated fashion (e.g. through a process that monitors CPU load and automatically adds more CPU resources once a usage threshold is reached).
- Virtualization: this technology enables the sharing of computer hardware amongst multiple users in a common resource pool, thanks to the abstraction of physical servers and of storage devices. This is, in particular, advantageous when using heterogeneous computing resources which need to be presented as one single platform or when different operating systems are to be run on the same hardware.

**2.1. Service models.** Cloud services can be classified into three main models:

- *Infrastructure as a Service* (IaaS): this service model is offered in a computing infrastructure that includes servers (typically virtualized) with specific computational capability and/or storage. The user controls all the storage resources, operating systems and applications deployed to, while he/she has limited control over the network settings. An example is Amazon's Elastic Compute Cloud (EC2), which allows the user to create virtual machines and manage them, and Amazon Simple Storage Service (S3), which allows to store and access data, through a web-service interface.
- *Platform as a Service* (PaaS): it allows the development, installation and execution of user-developed applications. Applications must be created using programming languages, libraries, services and tools supported by the provider that constitute the development platform provided as a service. An example is Google Apps Engine, which allows to develop applications in Java and Python, provides for both languages the SDK and uses a plugin for the Eclipse development environment.
- *Software as a Service* (SaaS): customers can use the applications provided by the cloud provider infrastructure. The applications are accessible through a specific interface. Customers do not manage the cloud infrastructure or network components, servers, operating systems or storage. In some cases, it is possible to manage specific configurations of the application.

**2.2. Delivery models.** Cloud services can be made available to users in different ways. In the following, a brief description of the delivery models is presented:

1. *Public Cloud*: public cloud services are offered by vendors who provide the users/customers the hardware and software resources of their data centers. Examples of public clouds are Amazon<sup>1</sup> (Amazon EC2 provides computational services, while Amazon S3 storage services); Google Apps (which provides software services like Gmail, Google Docs or Google Calendar, and development platform like Google App Engine); and Microsoft Azure<sup>2</sup>, that is a cloud computing platform and infrastructure, created by Microsoft, for building, deploying and managing applications and services through a global network of Microsoft-managed datacenters. It provides both PaaS and IaaS services and supports many different programming languages, tools and frameworks, including both Microsoft-specific and third-party software and systems.
2. *Private Cloud*: private cloud is configured by a user or by an organization for their exclusive use. Services are supplied by computers that are in the domain of the organization. To install a private cloud, several commercial and free tools are available (e.g. OpenStack<sup>3</sup>, Eucalyptus<sup>4</sup>, Open Nebula<sup>5</sup>, Terracotta<sup>6</sup> and VMWare Cloud<sup>7</sup>).
3. *Community Cloud*: it is an infrastructure on which are installed cloud services shared by a community or by a set of individuals, companies and organizations that share a common purpose and that have the same needs. The cloud can be managed by the community itself or by a third-party (typically a cloud service provider).
4. *Hybrid Cloud*: the cloud infrastructure is made up of two or more different clouds using different delivery models, which, while remaining separate entities, are connected by proprietary or standard technology that enables the portability of data and applications.

**3. Cloud Computing in Healthcare and Biomedicine.** Recent studies indicated that cloud computing can improve healthcare services and benefit biomedical research [10, 12, 20], by offering new possibilities.

An important driver for the adoption of cloud technology in the healthcare is to lower the cost of healthcare delivery. This cost has grown to such huge proportions that governments face serious funding issues. The recognition that technology can improve patient care while reducing costs has meant that governments are willing to push the traditionally slow healthcare industry to a faster pace of adoption. In [21], the authors pointed out that, based on Azures pricing model, the cost of their cloud-based telemedicine service is based on the use of CPU resources, which is charged by USD \$ 0.12 per hour, and the use of database, which is charged by USD \$9.99 per GB per month. Dudley and his colleagues [22] demonstrated that cloud computing is a viable and cheaper technology that enables large-scale integration and analysis for studies in genomic medicine. They compared computational and economic characteristics of a cloud-based service, with a local institutional cluster.

Another important driver is the big data growth in healthcare and bioinformatics [23]. As the amount of digital information increases, the ability to manage this data becomes a growing problem. This data holds the keys to future clinical advances, but often remains inaccessible to researchers. Cloud Computing can be the enabling factor for data sharing and integration at a large scale. In the paper [24], high performance computing (HPC) solutions in bioinformatics, Big Data analysis paradigms for computational biology, and the issues that are still open in the biomedical and healthcare fields are discussed. Specifically, the authors pointed out that cloud computing addresses Big Data storage and analysis issues in many fields of bioinformatics, thanks to the virtualization that avoid to move too big data.

Moreover, in the specific field of telemedicine, in order to support efficient monitoring and automated analysis of larger patient populations, it is essential to have an infrastructure that provides high throughput, high volume storage and reliable communication. *Horizontal scalability* (i.e. the ability for a system to easily expand

<sup>1</sup><https://aws.amazon.com>

<sup>2</sup><https://azure.microsoft.com>

<sup>3</sup><https://www.openstack.org>

<sup>4</sup><https://www.eucalyptus.com>

<sup>5</sup><http://opennebula.org>

<sup>6</sup><http://terracotta.org>

<sup>7</sup><http://www.vmware.com/>

its resource pool to accomodate heavier load) and *geographic scalability* (i.e. the ability to maintain performance, usefulness or usability regardless of expansion from concentration in a local area to a more distributed geographic pattern) represent two requirements that cloud computing can fulfill [25].

In the field of medical imaging, thanks to high-resolution imaging instruments, the volume of data will reach petabytes. Therefore, it is evident that the cloud computing model will provide a potential contribution to satisfy computational needs related to the reconstruction and analysis of medical images and to allow a wide sharing of imaging data, as well as advanced remote analysis.

In the following sections, the main cloud-based applications proposed in the fields of bioinformatics, molecular modeling, medical imaging, telemedicine and medical informatics are illustrated and discussed.

**3.1. Cloud-based Bioinformatics Solutions.** The cloud computing represents a cost-effective solution for the problems of storing and processing data in the context of bioinformatics. Thanks to the progress of high-throughput sequencing technologies, there was an exponential growth of biological data. Therefore, classical computational infrastructure for data processing have become ineffective and difficult to maintain [26, 27].

The traditional bioinformatics analysis involves downloading of public datasets (e.g. NCBI, Ensembl), installing software locally and analysis in-house. By entering the data and software in the cloud and providing them as a service, it is possible to get a level of integration that improves the analysis and the storage of bioinformatics big-data. In particular, as a result of this unprecedented growth of data, the provision of data as a service (*Data as a Service*, DaaS) is of extreme importance. DaaS provides a data storage in a dynamic virtual space hosted by the cloud and allows to have updated data that are accessible from a wide range of connected devices on the web. An example is represented by the DaaS of Amazon Web Services (AWS) [28], which provides a centralized repository of public data sets, including archives of GenBank, Ensembl, 1000 Genomes Project, Unigene, Influenza Virus. AWS public dataset<sup>8</sup> contains not only data related to the field of biology but also to other scientific fields, such as astronomy, chemistry, climate, economy, etc. All datasets are provided as public AWS services and so can be easily integrated into cloud-based applications.

In the following subsections, examples of SaaS, PaaS and IaaS for several tasks in bioinformatics domain are presented.

**3.1.1. Bioinformatics Tools deployed as SaaS.** In recent years, there have been several efforts to develop cloud-based tools to execute different bioinformatics tasks [29], e.g. genomic applications<sup>9 10</sup> [30, 31], sequences alignment [32], gene expression analysis<sup>11 12 13</sup> [33, 34, 35]. Other examples of SaaS bioinformatics tools are reported in the following.

**CloudBurst**<sup>14</sup> is a new parallel read-mapping algorithm optimized for mapping next-generation sequence (NGS) data to the human genome and other reference genomes, for use in a variety of biological analyses including SNP discovery, genotyping and personal genomics [36]. CloudBurst uses the open-source Hadoop implementation of MapReduce to parallelize execution using multiple compute nodes.

**PeakRanger**<sup>15</sup> [37] is a software package for the analysis in Chromatine ImmunoPrecipitation Sequencing (ChIP-seq) technique. This technique is related to NGS and allows to investigate the interactions between proteins and DNA. Specifically, PeakRanger is a peak caller software package that can be run in a parallel cloud computing environment to obtain extremely high performance on very large data sets.

**VAT** (Variant Annotation Tool)<sup>16</sup> [38] has been developed to functionally annotate variants from multiple personal genomes at the transcript level as well as to obtain summary statistics across genes and individuals. VAT also allows visualization of the effects of different variants, integrates allele frequencies and genotype data from the underlying individuals and facilitates comparative analysis between different groups of individuals.

<sup>8</sup><http://aws.amazon.com/publicdatasets>

<sup>9</sup><http://cloudaligner.sourceforge.net>

<sup>10</sup><http://bowtie-bio.sf.net/myrna>

<sup>11</sup><http://fx.gmi.ac.k>

<sup>12</sup><http://bowtie-bio.sourceforge.net/myrna/index.shtml>

<sup>13</sup><http://tinyurl.com/yumbedownload>

<sup>14</sup><http://sourceforge.net/projects/cloudburst-bio/>

<sup>15</sup><http://www.modencode.org/software/ranger/>

<sup>16</sup><http://vat.gersteinlab.org>

VAT can either be run through a command-line interface or as a web application. Finally, in order to enable on-demand access and to minimize unnecessary transfers of large data files, VAT can be run as a virtual machine in a cloud-computing environment.

**STORMSeq** (Scalable Tools for Open-source Read Mapping) [39], is a graphical interface cloud computing solution that performs read mapping, read cleaning and variant calling and annotation with personal genome data. At present, STORMSeq costs approximately 2 dollars and 510 hours to process a full exome sequence and 30 dollars and 38 days to process a whole genome sequence. The authors provide this open-access and open source resource<sup>17</sup> as a user-friendly interface in Amazon EC2.

In [40], the authors propose an efficient Cloud-based Epistasis cOmputing (**eCEO**) model for large-scale epistatic interaction in genome-wide association study (GWAS). Given a large number of combinations of SNPs (Single-nucleotide polymorphism), eCEO model is able to distribute them to balance the load across the processing nodes. Moreover, eCEO model can efficiently process each combination of SNPs to determine the significance of its association with the phenotype. Authors have implemented and evaluated eCEO model on their own cluster of more than 40 nodes. The experiment results demonstrate that the eCEO model is computationally efficient, flexible, scalable and practical. In addition, authors have also deployed the eCEO model on the Amazon Elastic Compute Cloud<sup>18</sup>.

**Cloud4SNP** [41] is a novel Cloud-based bioinformatics tool for the parallel preprocessing and statistical analysis of pharmacogenomics SNP DMET microarray data. It is a Cloud-based version of DMET- Analyzer [42], that has been implemented on the Cloud using the Data Mining Cloud Framework [43], a software environment for the design and execution of knowledge discovery workflows on the Cloud [44]. It allows to statistically test the significance of the presence of SNPs in two classes of samples using the well known Fisher test. Cloud4SNP uses data parallelism and employs an optimization technique to avoid the execution of useless Fisher tests, through the filtering of probes with similar SNPs distributions.

**ProteoCloud** [45] is a freely available, full-featured cloud-based platform to perform computationally intensive, exhaustive searches using five different peptide identification algorithms. ProteoCloud<sup>19</sup> is entirely open source, and is built around an easy to use and cross-platform software client with a rich graphical user interface. This client allows full control of the number of cloud instances to initiate and of the spectra to assign for identification. It also enables the user to track progress, and to visualize and interpret the results in detail. Table 3.1 lists the main SaaS bioinformatics applications and their essential features.

**3.1.2. Bioinformatics Platforms deployed as PaaS.** Currently, the most used platform (PaaS) for bioinformatics applications is **Galaxy Cloud**<sup>20</sup>, which is a Galaxy cloud-based platform for the analysis of data at a large scale. It allows anyone to run a private Galaxy installation on the Cloud exactly replicating functionality of the main site<sup>21</sup>, but without the need to share computing resources with other users. With Galaxy Cloud, unlike software service solutions, the users can customize their deployment as well as retain complete control over their instances and associated data; the analysis can also be moved to other cloud providers or local resources, avoiding concerns about dependence on a single vendor. Currently, a public Galaxy Cloud deployment is provided on the popular Amazon Web Services (AWS) cloud; however, it is compatible with Eucalyptus and other clouds [46]. **CloudMan**<sup>22</sup> [47] enables individual bioinformatics researchers to easily deploy, customize, and share their entire cloud analysis environment, including data, tools, and configurations.

In [48], a modular and scalable framework called **Eoulsan**<sup>23</sup>, based on the Hadoop implementation of the MapReduce algorithm dedicated to high-throughput sequencing data analysis, is presented. Eoulsan allows users to easily set up a cloud computing cluster and automate the analysis of several samples at once using various software solutions available. Tests with Amazon Web Services demonstrated that the computation cost is linear with the number of instances booked, as the running time with the increasing amounts of data. Eoulsan is implemented in Java, supported on Linux systems and distributed under the LGPL License.

<sup>17</sup><http://www.stormseq.org>

<sup>18</sup><http://www.comp.nus.edu.sg/wangzk/eCEO.html>

<sup>19</sup><http://proteocloud.googlecode.com>

<sup>20</sup><http://galaxy.psu.edu>

<sup>21</sup>Galaxy is an open source, web-based platform for data intensive biomedical research; <http://usegalaxy.org/>

<sup>22</sup><http://cloudman.irb.hr>

<sup>23</sup><http://transcriptome.ens.fr/eoulsan/>

TABLE 3.1  
SaaS Bioinformatics Applications

Name	Available	Domain	Task/Data	Cloud Type
Cloud Aligner [30]	YES	Genomics (Genome resquencing, short-read aligner)	A map/reduce based application for mapping short reads generated by the next-generation sequencing machine	Public Cloud (AWS)
CloudBlast [32]	YES	Genomics	Massive sequence alignment	Public Cloud (AWS)
Fx [33]	YES	Genomics	RNA-Seq analysis tool	Public Cloud (AWS)
Myrna [34]	YES	Genomics (RNA-sequencing)	Calculation of differential gene expression in large RNA-seq datasets	Public Cloud (AWS)
YunBe [35]	YES	Transcriptomics	Gene set analysis tool for biomarker identification	Public Cloud (AWS)
CloudBurst [36]	YES	Genomics (Genome resquencing, short-read aligner)	Parallel Mapping of NGS data	Public Cloud (AWS)
PeakRanger [37]	YES	Genomics (Sequencing)	Peak caller for ChIP-Seq data	Binaries for an Amazon EC2 image will be available
VAT [38]	YES	Genomics	Framework to functionally annotate variants in personal genomes	Public Cloud (AWS)
StormSeq [39]	YES	Genomics (Sequencing)	Parallel Processing Personal Genomics Data	Public Cloud (AWS)
eCEO [40]	YES	Genomics (Sequencing)	Epistasis Computing Model of SNPs data	Public Cloud (AWS)
Cloud4SNP [41]	No	Genomic	Microarray data analysis	Private Cloud (Microsoft Azure)
ProteoCloud [45]	YES	Proteomics (Mass spectrometry)	Tools for identification of very large sets of spectra using five different peptide identification algorithms	Public Cloud (AWS)

The Table 3.2 summarizes the bioinformatics platforms (PaaS) introduced so far by evidencing the application domain, the type of cloud (public, private, etc.) and the information relative to possible availability.

**3.1.3. Bioinformatics Virtual Machines deployed as IaaS.** **Bionimbus**<sup>24</sup> [49] is an open source cloud-computing platform used by a variety of projects to process genomics and phenotypic data. It is based primarily upon OpenStack, which manages on-demand virtual machines that provide the required computational

<sup>24</sup><http://bionimbus.opensciencedatacloud.org>

TABLE 3.2  
*PaaS Bioinformatics Applications*

Name	Avail	Domain	Task	Cloud Type	O.S.	Programming Language
GalaxyCloud [46]	YES	Bioinformatics	Cloud-scale Galaxy for large-scale data analysis	Public Cloud (AWS)	Not available	Not available
CloudMan [47]	YES	Bioinformatics	Platform for tool, data, and analysis distribution.	Public Cloud (Nectar)	*NIX	Python
Euolsan [48]	YES	Genomics	High throughput sequencing analyses	Public Cloud (AWS)	Linux	Java

resources, and GlusterFS, which is a high-performance clustered file system. Bionimbus also includes Tukey, which is a portal, and associated middleware that provides a single entry point and a single sign on for the various Bionimbus resources; and Yates, which automates the installation, configuration, and maintenance of the software infrastructure required. Bionimbus is required to operate under a set of guidelines developed by NIST called FISMA. These requirements served as a guide to improve the security of Bionimbus and to generate the associated documentation. The authors modified the OpenFISMA application to help automate this process, e.g. with their modifications, security scans are performed automatically at regular intervals and the results recorded.

Cloud Virtual Resource, **CloVR**<sup>25</sup> [50] is a new desktop application for push-button automated sequence analysis that can utilize cloud computing resources. CloVR is implemented as a single portable virtual machine (VM) that provides several automated analysis pipelines for microbial genomics, whole genome and metagenome sequence analysis. In addition CloVR supports use of remote cloud computing resources to improve performance for large-scale sequence processing. Supported clouds include the commercial Amazon Elastic Compute Cloud and the academic platforms DIAG [51] and Magellan [52] To provide security and help ensure data privacy, each remote cluster of CloVR VMs uses a unique, randomly generated authentication key. This key is used to enable secure data transfer between instances with Secure Shell (SSH) both within the cloud and over the Internet and between the local client VM and master cloud CloVR VMs.

**Cloud BioLinux** [53]<sup>26</sup> is a publicly accessible Virtual Machine (VM) that enables scientists to quickly provision on-demand infrastructures for high-performance bioinformatics computing using cloud platforms. Users have instant access to a range of pre-configured command line and graphical software applications, including a full-featured desktop interface, documentation and over 135 bioinformatics packages for applications including sequence alignment, clustering, assembly, display, editing, and phylogeny. Cloud BioLinux is available through the Amazon EC2 cloud. For higher security, instead of providing a password through the EC2 launch instance wizard, a user can take the additional step of setting up Secure Shell (SSH) keys, which is fully described in the documentation. Besides the Amazon EC2 cloud, the authors started instances of Cloud BioLinux on a private Eucalyptus cloud and demonstrated access to the bioinformatics tools interface through a remote connection to EC2 instances from a local desktop computer.

**3.1.4. Cloud-based Solutions for Molecular Modeling.** Different molecular modeling tasks, such as molecular dynamics simulations, quantum mechanical calculations or 3D virtual library construction, are very

<sup>25</sup><http://clovr.org>

<sup>26</sup><http://cloudbiolinux.org>

TABLE 3.3  
*Bioinformatics IaaS*

Name	Avail.	Domain	Task	Cloud Type	Operating System	Program- ming Lan- guage
CloudBio-Linux [53]	YES	Bioinformatics	A publicly virtual machine for high performance bioinformatics computing	Public (AWS) and private Cloud (Eucalyptus)	Linux, Windows, Mac OSX	Python
CloVR [50]	YES	Genomics	A virtual machine for automated and portable sequence analysis	Cloud Mid-ware EC2, Eucalyptus, Nimbus	Platform independent	Not available
Bionimbus [49]	YES	Genomics	Cloud-based system for genomic data	Private cloud (OpenStack)	Not available	Not available

computationally intensive tasks and generate or process huge amount of data. Therefore, cloud computing model guarantees for these types of molecular modeling tasks scalability, reliability, and lower cost [54]. The use of cloud computing for molecular modeling applications is still in its infancy, because in order to use cloud computing, many technical tasks must be performed by the user such as configuring the compute nodes, installing the required software, and launching, monitoring and terminating the computation [55].

One of the most sophisticated SaaS for cloud computing are provided by commercial providers such as CycleCloud<sup>27</sup>, which offers clusters with preinstalled applications that are widely used in the area of bioinformatics, proteomics, and computational chemistry. The most interesting from a molecular modeling perspective are Gromacs, to perform molecular dynamics (MD) simulations, and ROCS to perform shape-based screening of compounds.

Recently announced AceCloud<sup>28</sup> is a commercial on-demand GPU cloud computing service. The optimization of MD simulation packages for running on GPUs, such as Amber, Gromacs, Lammps, and NAMD is reported to result in large performance increases compared to calculations run on CPUs [56].

The Rosetta software suite<sup>29</sup> facilitates molecular modeling tasks such as prediction of protein and RNA 3D structures, protein-ligand and protein-protein docking, antibody modeling, and enzyme design. Because of the large search spaces of these problems, finding good solutions is highly computationally demanding. To meet this challenge a Seattle biomedical software company Insilicos announced the formation of RosettaATCloud LLC, a joint venture with the University of Washington, that developed Rosetta@cloud, which offers the Rosetta software suite as a cloud-based service hosted on Amazon Web Services.

AutoDockCloud [57] is a workflow system that enables distributed screening on a cloud platform using the molecular docking program AutoDock. AutoDockCloud is based on the open source framework Apache Hadoop, which implements the MapReduce paradigm for distributed computing. In addition to the docking procedure itself, preparation steps, such as generating docking parameters and ligand input files, can be automated and distributed by AutoDockCloud.

<sup>27</sup><http://www.cyclecomputing.com/cyclecloud/applications>

<sup>28</sup><http://www.acellera.com/products/acecloud/>

<sup>29</sup><https://www.rosettacommons.org>

**3.2. Cloud-based Medical Informatics Solutions.** A major challenge for medical informatics concerns the storage of heterogeneous data into a single searchable database for clinical purposes or research. The cloud allows greater sharing and integration of Electronic Health Record (EHR), also allowing a reduction in costs related to hardware, software, network and staff. Following these considerations, it is expected a considerable increase in the use of the cloud for storing personal health information online [58].

Many studies have proposed different cloud-based frameworks in an attempt to improve the EHR. Among these, Chen et al. [59] have proposed a new system of access control in cloud computing environment that allows to securely access the patient's medical record even in multi-user system.

Doukas et al. [60] have presented the implementation of a mobile system that allows data storage of e-health, updating and recovery with cloud computing.

In addition to the works presented so far as result of academic research, there are also several commercial systems made by companies that provide services of medical records.

An example is the Microsoft HealthVault<sup>30</sup> platform developed by Microsoft for managing health information of its members. The system helps people to collect, store and share their medical information with family members and health care providers. In addition, it also allows the management of fitness and diet connecting with third-party applications.

HealthVault gives particular attention to privacy and security of health information users. In the following, some security mechanisms are listed:

- Authentication: users can access the system through various mechanisms such as Facebook, Windows Live ID;
- Authorization: The system asks the user's permission before enabling any data access by an application;
- User Control: Users are able to control how their data have to be shared, explicitly granting or revoking consent for third-party applications to access their account. Users are also able to close the account at any time or to remove specific information.
- Auditing: a log that records the data access is built-in into the system and available to users.
- Source of data: the data in HealthVault contain the source of the data and applications and one can use this information to determine how to treat information coming from different sources. The digital signature also allows users to optionally verify the integrity of the data and their source.

Fusion [61] is a cloud-based experimental platform that aims to manage in a simple and secure way, health information of patients allowing them to share between operators in the healthcare system. The key features are low cost and large scale. With regard to data protection, the Fusion architecture includes mechanisms to enable the security and privacy and traceability of access to stored medical data. Specifically, since all data residing in the cloud are encrypted, and the Fusion Store by itself cannot decrypt the data, the security and privacy of data are guaranteed. Different health care professionals can use Fusion to store patient information and share it securely. Fusion also supports the integration of proprietary systems of health information management and provides integrated data sharing mechanisms.

Several hospitals and healthcare organizations have been adopted cloud computing solution. On July 2011 Chelsea and Westminster Hospital set a cloud computing system to manage and store their EHRs. With this system, patients have full control over who has access to their health records. However the Cloud provider company must prove the security of its system. To solve this problem the Cloud provider company has implemented a security mechanism in which users have to pass multiple ID checkpoints to access the database [83].

Another example of Cloud Computing and e-Health services is found at the "Bambino Gesù" Italian Hospital. This hospital, placed in Rome, is famous for being one of the largest research and treatment centers in the field of pediatrics. Since hospital people are using Cloud, they have experienced advantages such as: better collaboration between the medical staff, better connection with patients and more free time for the IT group [62].

Cloudera has been working with Mount Sinai School of Medicine clinical and academic faculty to develop new methods and systems for analyzing data in a variety of important areas including genomics and multiscale biology.

---

<sup>30</sup><https://www.healthvault.com/>

Recently, Apple has announced new Health app that collects all health and fitness data (heart rate, calories burned, blood sugar, cholesterol) in order to give to the user a clear and current overview of his/her health. It is possible also to create an emergency card with important health information, for example, the blood type or allergies.

The Google project, called Baseline Study, will collect anonymous genetic and molecular information from 175 people – and later thousands more – to create what the company hopes will be the fullest picture of what a healthy human being should be. Baseline will amass a much larger and broader set of new data. The hope is that this will help researchers detect killers such as heart disease and cancer far earlier, pushing medicine more toward prevention rather than the treatment of illness. Google has already built one of the world’s largest networks of computers and data centers to serve online-search results quickly.

**3.3. Cloud-based Medical Imaging Solutions.** Many studies have demonstrated the utility of MapReduce to solve the problems of medical imaging on a large scale in a cloud computing environment. For example, Meng et al. [63] have developed a fast and scalable technique for image reconstruction of 4D CT (Four Dimension Computer Tomography) cone beam with MapReduce in a cloud computing environment.

Avila-Garcia et al. [64] proposed a framework based on cloud for the analysis of images in order to diagnose colorectal cancer.

Silva et al. [65] proposed a system based on a set of DICOM routers interconnected through an infrastructure of public cloud to support the exchange of medical images between institutions.

In the field of radiology, Patel [66] highlighted the main advantages of the use of cloud-based SaaS, PaaS and IaaS applications. Specifically, SaaS allows access to three-dimensional visualization and quantitative analysis tools from any computer with a web client and internet access. This permits to solve license problems and the need of powerful local workstation. iNtuition UNLIMITED<sup>31</sup> cloud-based service by TeraRecon, Inc is an example of commercial product for radiologists. Moreover, the sharing of images between radiologist of different institutions and the increasing storage capacities needed bring to adoption of cloud-based solutions. Commercial vendors such as Insite One and Pacs Drive offer cloud based imaging archiving and storing solutions.

Security and privacy of patients images should be ensured in order to avoid unauthorised access. Kagadis et al. [67] discuss the need to adopt specific methods of ensuring data security including encrypting all data transferred over the internet using SSL (Secure Sockets Layer), and giving only an authorized user access to the system.

**3.4. Cloud-based Telemedicine Solutions.** In many countries, increasing longevity and declining fertility rates have led to the aging of the population and, consequently, the need for medical care is increased. In addition, there is a large population that does not receive good health care services because they live in rural communities.

Telemedicine, through the provision of remote monitoring services and medical assistance, allows to reach people living in rural areas and to reduce the healthcare costs. Moreover, the use of cloud computing could allow to delivery better telemedicine services. Specifically, the cloud can help provide effective treatment and care of patients due to its benefits such as on-demand access anywhere, anytime, low costs and high elasticity.

In this subsection, a description of cloud-based telemedicine applications are illustrated, highlighting the advantages and limits of the proposed approaches. Table 3.4 listed the features of the cited telemedicine applications. The major number of reported studies focus on telecardiology [68].

Pandey et al. in [25] propose the development of an autonomic cloud environment that collects people’s health data and stores them to a cloud-based information repository and facilitates analysis on the data using software services hosted in the cloud. The scheme by Pandey et al. integrates mobile and Cloud technologies with electrocardiogram (ECG) sensors to enable remote monitoring of patients with heart-related problems such as cardiac arrhythmias. The patients connect the sensors to their body and then run an app on a mobile device. The app connects to the sensors via Bluetooth. The app will then periodically upload data to the Cloud. The user can then download graphs from the Cloud which represent the users health states. The scheme also implements middleware in the Cloud. There are web services for users to analyse their ECG, draw graphs, etc. The current scheme does not handle security issues, even if the authors propose and discuss the use of a trusted

---

<sup>31</sup><http://www.terarecon.com/cloud/>

third party-certification authority to implement a Public Key Infrastructure (PKI) based authentication and authorization system.

Also in the paper [69], security and privacy issues are not discussed. The authors propose a cloud-based system for monitoring and real-time analysis of electrocardiographic signals (ECG). The system has been designed to be accessible from anywhere using a smartphone, a tablet, desktop or laptop. The ECG signals can be collected by wearable ECG devices and transmitted via Bluetooth to the client, which then transmits to the cloud. In particular, the web application has been developed using the Spring framework<sup>32</sup>. The final application was deployed to a server in the Amazon Cloud using Amazon Web Services (AWS). The transmission of data between the client and the server has been done by using HTTPS (Hypertext Transfer Protocol Secure). The server, upon receiving the data, controls the quality of the signals and, if necessary, applies appropriate pre-processing and enhancement techniques. Subsequently, the system extracts information related to heart rate variability (RR interval<sup>33</sup>). The system provides an interface for the patient and for the physician. The patient may display the ECG trace and the extracted parameters. The physician can visualize these data, and according to their value, he/she could make a diagnosis that is sent to the patient.

In [21], Hsieh et al. describe the implementation of a cloud computing based telemedicine application, which can upload 12-lead ECG reading to a cloud service from which they can be visualized on a variety of network connected devices, both mobile and fixed. This cloud application supports a variety of clinically used 12-lead ECG devices, such as Philips XML-ECG, HP compatible SCP-ECG, and Mortara DICOM-ECG by first transcoding the data from a vendor-specific format to an XML format in a public cloud computing platform, Microsoft Azure. To safeguard the ECG data in the cloud, they have designed a security mechanism to prevent them from theft and protect the patients privacy, including authentication for the use of Web roles and Worker roles, data encryption during message communications among roles, and ECG file encryption and verification while ECG report are retrieved in storage account and database. Moreover, ECG files are transmitted via secure sockets layer (SSL) based HTTP (HTTPS) where ECG files are protected by certificate based encryption and verification instead of plain text based HTTP.

Body Sensor Network (BSN), with their huge amount of gathered sensor data and their limited processing power, can be empowered by exploiting a Cloud computing infrastructure to realize an integrated platform that provides [70]:

- the ability to utilize heterogeneous sensors;
- scalability of data storage;
- scalability of processing power for different kinds of data analysis;
- global access to the processing and storage infrastructure;
- easy sharing of results;
- pay- as-you-go pricing for using BSN services;

Fortino [71] introduced the BodyCloud architecture which enables the management and monitoring of body sensor data via the Cloud. It provides functionality for receiving and managing sensor data in a seamless way from a body sensor network (BSN). BodyCloud also comprises of a scalable framework that allows support for multiple data streams required for running concurrent applications. Many challenges (interoperability, heterogeneity, security, data validation and consistency) strictly related to BSNs are discussed in [72].

In [73] a cloud computing based Voice over IP (VoIP) service for diabetic patients self care management, is presented. The patients subscribing the cloud service received VoIP calls with pre-recorded voice messages as self-care reminders. In this study, the participants obtained better glycemic control than the patients without subscribing the service.

In [74] a cloud-based intelligent system for real-time monitoring of users with diabetes (Cloud Based Intelligent Health Care Service, CBIHCS), has been presented. CBIHCS allows the monitoring of different vital parameters such as blood pressure, glucose and ECG. The collected data are sent via bluetooth to mobile devices which then transmit them to the cloud. Measured signals and patients information are stored on the cloud. Considering the analysis of the data, a PCA (Principal Component Analysis) is applied for the selection

<sup>32</sup><http://projects.spring.io/spring-framework/>

<sup>33</sup>the time elapsing between two consecutive R waves in the electrocardiogram or the interval from the peak of one QRS complex to the peak of the next as shown in the electrocardiogram

of attributes, and classification techniques, such as KNN and Naïve Bayes have been used to determine the state of health of the users. Furthermore, the authors have implemented statistical prediction techniques to determine the patterns of resource use and have proposed a simple heuristic for a dynamic elastic infrastructure. In particular, the authors highlight that in the development of the system, the standard Web Service Resource Framework (WSRF) has been adopted in order to facilitate the deployment of the final system on a cloud, like Amazon EC2. They carried out two types of experiments: the aim of the first experimentation was to assess the classifier's performances and to check the proper functioning of the system; in the second experimentation a study on dynamic allocation of resources was performed. Moreover the authors have been addressed security challenges in a cloud hosted health care application: they have implemented security mechanisms at multiple levels and provide role based access control to ensure the protection of critical medical data of patients. They propose to integrate the use of symmetric cryptosystems for authentication and role based access control (RBAC) mechanisms for authorization.

In Kim et al. [75], another cloud-based system for vital signs monitoring is presented. The framework is composed of four distinct modules: a module to receive vital sign data using standardized messaging methods; a module to transform these data into a standardized schema; a module to evaluate the health state using biosignal data; and a service-oriented architecture (SOA) component module that allows users to access medical services with standardized messaging methods. Security and privacy issues have not been discussed, but a new concept of HaaS Healthcare as Service has been introduced: healthcare systems are defined as services in order to secure interoperability.

The paper [76] presents an expert diagnosis system based on cloud computing. It classifies a users fitness level based on supervised machine learning techniques. This system is able to learn and make customized diagnoses according to the users physiological data, such as age, gender, and body mass index (BMI). In addition, an elastic algorithm based on Poisson distribution is presented to allocate computation resources dynamically. The paper does not present any detailed information about cloud implementation.

In [77], a new mobile app, Dental Calendar, combined with cloud services specific for dental care, has been created. This new system would remind patients about every scheduled apointment, and help them take pictures of thier own oral cavity parts that require dental treatment and send them to dentists along with a symptom description. All this information would be sent to dentists through cloud services. The authors propose a general system's architecture, but they do not discuss it in detail.

Finally, Thilakanathan et al. [78] addressed the issues of privacy and security in the domain of mobile telecare and Cloud computing. They demonstrated a telecare application that will allow doctors to remotely monitor patients via the Cloud. They use this system as a basis to showcase a model that will allow patients to share their health information with other doctors, nurses or medical professionals in a secure and confidential manner. The key features of the model include the ability to handle large data sizes and efficient user revocation. The authors also addressed the problems of achieving efficient user revocation, especially when considering large data sizes. For this, they defined a secure data sharing model and protocol, and demonstrated the feasibility of the protocol through a prototype which combines smartphone, Bluetooth and Cloud computing technologies.

**4. Cloud challenges in healthcare and biomedicine.** Clinical data must be hosted on publicly accessible servers according to privacy and security rules, such as the Health Insurance Portability and Accountability Act (HIPAA). The EHRs are exposed to possible abuse and require security measures based on the identity management, access control, policy integration and compliance management.

Typical entities in a cloud based health record system are patients, hospital staff, such as doctors, nurses, pharmacies, and laboratory staff, insurance companies, and the cloud service providers. Due to the distributed architecture of the cloud, the patient EHRs are stored at and shared among many third-party providers. Therefore, the data is susceptible to unauthorized access and attacks. Specifically, the paper [79] claims that storing huge volumes of patients' sensitive medical data in third-party cloud storage is susceptible to loss, leakage or theft. The privacy risk of cloud environment includes the failure of mechanisms for separating storage, memory, routing, and even reputation between different tenants of the shared infrastructure. The centralized storage and shared tenancy of physical storage space means the cloud users are at higher risk of disclosure of their sensitive data to unwanted parties.

Threats to the data privacy in the cloud include spoofing identity, tampering with the data, repudiation,

TABLE 3.4  
*Cloud-based telemedicine systems*

References	Application main	Do-	Task/Data	Cloud	Security and Privacy
[25]	Telecardiology		ECG monitoring	Public Cloud (Amazon Web Services)	No
[69]	Telecardiology		ECG real-time monitoring	Public Cloud (Amazon Web Services)	No
[21]	Telecardiology		ECG monitoring	Public Cloud (Microsoft Azure)	Yes
[71]	Body Sensor Network (BSN)	Net-	Management and Monitoring of body sensor data	Not available	Yes
[73]	Teleconsulting		Diabetes	Not available	No
[74]	Diabetes		Monitoring of different vital parameters	Public Cloud (Amazon Web Services)	Yes
[75]	Telemonitoring		VMonitoring of vital parameters	Not available	Not available
[76]	Wellness		Expert system for health fitness level classification	Not available	Not available
[77]	Dental care		Appointment reminders and transmission of oral cavity pictures	Not available	Not available

and information disclosure. In spoofing identity attack, the attacker pretends to be a valid user whereas data tampering involves malicious alterations and modification of the content. Repudiation threats are concerned with the users who deny after performing an activity with the data. Information disclosure is the exposure of information to the entities having no right to access information. The same threats prevail for the health data stored and transmitted on the third-party cloud servers.

Therefore, confidentiality and integrity of the stored health data are the most important challenges elevated by the healthcare and biomedicine cloud-based systems.

A secure protection scheme will be necessary to protect the sensitive information of the medical record. There is considerable work on protecting data from privacy and security attacks. NIST [80] has developed guidelines to help consumers to protect their data in the Cloud. The paper [81] evidences that using cryptographic storage significantly enhances security of the data. The paper discusses the main mechanisms to be adopted in order to guarantee and satisfy the previous cited issues. Specifically, the authors present and discuss the utility of cryptographic and non-cryptographic approaches. The cryptographic approaches to mitigate the privacy risks utilize certain encryption schemes and cryptographic primitives. Conversely, non-cryptographic approaches mainly use a policy based authorization infrastructures that allow the data objects to have access control policies. Particularly, in the public cloud environment operated by the commercial service providers and shared by several other customers, data privacy and security are the most important requirements.

Abbas et al. [82] summarized the security and privacy requirements for cloud-based applications in healthcare and biomedicine fields in the following way:

1. Integrity: it is needed to ensure that the health data captured by a system or provided to any entity is true representation of the intended information and has not been modified in any way.
2. Confidentiality: the health data of patients is kept completely undisclosed to the unauthorized entities.
3. Authenticity: the entity requesting access is authentic. In the healthcare systems, the information provided by the healthcare providers and the identities of the entities using such information must be

verified.

4. **Accountability:** an obligation to be responsible in light of the agreed upon expectations. The patients or the entities nominated by the patients should monitor the use of their health information, whenever that is accessed at hospitals, pharmacies, insurance companies etc.
5. **Audit:** it is needed to ensure that all the healthcare data is secure and all the data access activities in the e-Health cloud are being monitored.
6. **Non-Repudiation:** repudiation threats are concerned with the users who deny after performing an activity with the data. For instance, in the healthcare scenario neither the patients nor the doctors can deny after misappropriating the health data.
7. **Anonymity:** it refers to the state where a particular subject cannot be identified. For instance, identities of the patients can be made anonymous when they store their health data on the cloud so that the cloud servers could not learn about the identity.
8. **Unlinkability:** it refers to the use of resources or items of interest multiple times by a user without other users or subjects being able to interlink the usage of these resources. More specifically, the information obtained from different flows of the health data should not be sufficient to establish linkability by the unauthorized entities.

Finally, in the cloud, physical storages could be widely distributed across multiple jurisdictions, each of which may have different laws regarding data security, privacy, usage, and intellectual property. For example, the US Health Insurance Portability and Accountability Act (HIPAA) restricts companies from disclosing personal health data to nonaffiliated third parties. Similarly, the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) limits the powers of organizations to collect, use, or disclose personal information in the course of commercial activities. However, a provider may, without notice to a user, move the users information from jurisdiction to jurisdiction. Data in the cloud may have more than one legal location at the same time, with differing legal consequences.

In summary, security and privacy of data are the main challenges to be faced when deploying healthcare and biomedicine applications on the cloud. However, while in the healthcare community this awareness starts to be recognized by the various actors involved in data production, collection and decision making, as demonstrated by several initiatives and approaches developed, in biomedicine applications, based on the integrated analysis of biological (mainly omics) data and clinical data, there is not yet a clear indication on how security and privacy issues are to be faced. Most probably, the solutions that start to be adopted in healthcare could be adapted to biological and clinical data used in such applications.

**5. Conclusions.** Applications and services in health care and biomedicine pose quite demanding requirements. The fulfillment of these requirements could results in an improvement in the provision of services to patients in health care, as well as an increase in knowledge in the biomedical field. As shown and discussed in this paper, in order to meet these requirements, the use of the cloud computing model is required. Although supercomputing or Grid Computing can provide the computational power and the storage required in biomedical applications such as medical imaging or electronic medical record, the elasticity in providing such resources, a clear definition of Service Level Agreement, and the possibility for the customer to use a pay-per-use model, make the Cloud more suitable to support those applications.

Naturally, the adoption of this technology with its benefits will determine a reduction of costs and the possibility of also providing new services. Different comparative analysis demonstrate the potentials of cloud technology in reducing cost of IT organization: in the case of cloud adoption, institutions and laboratories are free from the expense and having to install and maintain applications locally [83]. Knaus et al. [84] focus on the financial aspects of grid and cloud computing. The authors discuss and compare the costs of a local cluster maintained by a department to a comparable cloud offer. Specifically, they list all direct costs and estimate all indirect costs and, with the help of a simulation study they estimate the costs for typical research projects in biomedicine. In their case, they conclude that using a cloud is a more viable option.

Also in the paper [85], the authors analyzed some economic and practical aspects of exploiting cloud computing for the in silico drug discovery and they highlight the advantages for small-medium laboratories working in the field of biotechnology, which typically do not have the possibility to invest a sufficient amount of time and money in creating and maintaining an in-house ICT infrastructure that suits the processing of the

large amount of bioinformatics data that are nowadays produced.

However, it is important to emphasize that the use of cloud in these fields is featured still by a number of open issues, such as the security and privacy, that require a rapid and efficient solution.

**6. Acknowledgments.** This work has been partially funded by the *PON DICET- INMOTO-ORCHESTRA (PON04a2-D)* project, funded by the Italian Ministry of Research and Education (MIUR).

#### REFERENCES

- [1] P. MELL AND T. GRANCE, *The NIST definition of cloud computing*, 2011.
- [2] M. ARMBRUST, A. FOX, R. GRIFFITH, A. D. JOSEPH, R. KATZ, A. KONWINSKI, G. LEE, D. PATTERSON, A. RABKIN, I. STOICA AND M. ZAHARIA, *A view of cloud computing*, Commun ACM, 53(4):50-58, 2010.
- [3] A.M.H. KUO, *Opportunities and Challenges of Cloud Computing to Improve Health Care Services*, J Med Internet Res., 13(3):e67, 2013.
- [4] E.J.SCHWEITZER, *Reconciliation of the cloud computing model with US federal electronic health record regulations*, J Am Med Inform Assoc., 2011, doi: 10.1136/amiajnl-2011-000162.amiajnl-2011-000162.
- [5] J.HAUGHTON, *Year of the underdog: Cloud-based EHRs*, Health Manag Technol,32(1):9, 2011.
- [6] J. KABACHINSKI, *What's the forecast for cloud computing in healthcare?*, Biomed Instrum Technol., 45(2):14650, 2011.
- [7] Y. LIN, C. YU AND Y. LIN, *Enabling Large-Scale Biomedical Analysis in the Cloud*, BioMed Research International, 2013, <http://dx.doi.org/10.1155/2013/185679>.
- [8] C. LIN, S. S ABDUL, D. L. CLINCIU, J. SCHOLL, X. JIN, H. LU, S. S. CHEN, U. IQBAL, M. J. HEINECK AND Y. LI, *Empowering village doctors and enhancing rural healthcare using cloud computing in a rural area of mainland China*, Computer methods and programs in biomedicine, 113:585592, 2014.
- [9] J. EKANAYAKE, T. GUNARATHNE AND J. QIU, *Cloud Technologies for Bioinformatics Applications*, IEEE Transactions on parallel and distributed systems, 22(6):998-1011, 2011.
- [10] S. P. AHUJA, S. MANI AND J. ZAMBRANO *A Survey of the State of Cloud Computing in Healthcare*, Network and Communication Technologies, 1(2):12–19, 2012.
- [11] M. CANNATARO, *Computational Grid Technologies for Life Sciences, Biomedicine and Healthcare*, Medical Information Science Reference Hershey, PA: IGI Global Press; 2009.
- [12] M. J. A. EUGSTER, M. SCHMID, H. BINDER AND M. SCHMIDBERGER, *Grid and Cloud Computing Methods in Biomedical Research*, Methods of Information in Medicine 1:62–64, 2013.
- [13] R. BUYYA, C. S. YEO, S. VENUGOPAL, J. BROBERG AND I. BRANDIC, *Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility*, Future Generation Computer Systems, 25:599–616, 2009.
- [14] R. L. GROSSMAN, *The case for cloud computing*, IT Professionals, 11:23–27, 2009.
- [15] L. M. VAQUERO, L. RODERO-MERINO, J. CACERES AND M. LINDNER, *A break in the clouds: Towards a cloud definition*, ACM SIGCOMM Computer Communication Review, 39:50–55, 2009.
- [16] S. MARSTON, Z. LI, S. BANDYOPADHYAY, J. ZHANG AND A. GHALSASI, *Cloud computing - the business perspective*, Decision Support Systems, 51:176-189, 2011.
- [17] INTEL, *Intel's Vision of the Ongoing Shift to Cloud Computing*, (2010), <http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/cloud-computing-intel-cloud-2015-vision.pdf>.
- [18] R. HILL, L. HIRSCH, P. LAKE AND S. MOSHIRI, *Guide to Cloud Computing*, 2013.
- [19] OGF PRODUCTION GRID INFRASTRUCTURE, *GFD.181 Glossary of Acronyms and Terms*, 20 March 2011.
- [20] A. ROSENTHAL, P. MORK, M. H. LI, J. STANFORD, D. KOESTER AND P. REYNOLDS, *Cloud computing: A new business paradigm for biomedical information sharing*, Journal of Biomedical Informatics, 43:342–353, 2010.
- [21] J. HSIEH AND M. W. HSU, *A cloud computing based 12-lead ECG telemedicine service*, BMC Medical Informatics and Decision Making, 12:77, 2012.
- [22] J. T. DUDLEY, Y. POULIOT, J. R. CHEN, A. A. MORGAN AND A. J. BUTTE, *Translational Bioinformatics in the cloud: an affordable alternative*, Genome Medicine, 2(51), 2010.
- [23] GREENE C. S, TAN J., UNG M., MOORE J. H. AND CHENG C., *Big data bioinformatics*, Journal of Cell Physiology 229(12):1896-1900, 2014.
- [24] I. MERELLI, H. PREZ-SNCHEZ, S. GESING AND D. DAGOSTINO, *Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives*, BioMed Research International, 2014, <http://dx.doi.org/10.1155/2014/134023>.
- [25] S. PANDEY, W. VOORSLUYS, S. NIU, A. DOKER AND R. BUYYA, *An autonomic cloud environment for hosting ECG data analysis services*, Future Generation Computer Systems, 28:147–154, 2012.
- [26] E. E SCHADT, M. D. LINDERMAN, J. SORENSON, L. LEE AND G. P. NOLAN, *Cloud and heterogeneous computing solutions exist today for the emerging big data problems in biology*, Nature Reviews Genetics, 12(3):224, 2011.
- [27] R. L. GROSSMANN AND K. P. WHITE, *A vision for a biomedical cloud*, Journal of Internal Medicine 271(2):122–130, 2011.
- [28] V. A. FUSARO, P. PATIL, E. GAFNI, D. P. WALL AND P. J. TONELLATO, *Biomedical Cloud Computing With Amazon Web Services*, Plos Computational Biology, 7(8), 2011.
- [29] L. DAI, X. GAO, Y. GUO, J. XIAO AND Z. ZHANG, *Bioinformatics clouds for big data manipulation*, Biology Direct, 7(43), 2012.

- [30] T. NGUYEN, W. SHI AND D. RUDEN, *CloudAligner: A fast and full-featured MapReduce based tool for sequence mapping*, BMC Research Notes, 4(171), 2011.
- [31] B. LANGMEAD, M. C. SCHATZ, J. LIN, M. POP AND S. L. SALZBERG, *Searching for SNPs with cloud computing*, Genome Biology, 10:R134, 2009.
- [32] A. MATSUNAGA, M. TSUGAWA AND J. FORTES, *CloudBLAST: Combining MapReduce and Virtualization on Distributed Resources for Bioinformatics Applications*, eScience, 2008. eScience '08.
- [33] D. HONG, *FX: an RNA-Seq analysis tool on the cloud*, Bioinformatics, 28(5):721–723, 2012.
- [34] B. LANGMEAD, K. D. HANSEN AND J. T. LEEK, *Cloud-scale RNA-sequencing differential expression analysis with Myrna*, Genome Biology 11(R83), 2010.
- [35] L. ZHANG, S. GU, B. WANG, Y. LIU, Y. AND F. AZUAJE, *Gene set analysis in the cloud*, Bioinformatics 28(2):294–295, 2012.
- [36] M. C. SCHATZ, *CloudBurst: highly sensitive read mapping with MapReduce*, Bioinformatics, 25(11):1363–1369, 2009.
- [37] X. FENG, R. GROSSMAN AND L. STEIN, *PeakRanger: A cloud-enabled peak caller for ChIP-seq data*, BMC Bioinformatics, 12(139), 2011.
- [38] L. HABEGGER, S. BALASUBRAMANIAN, D. Z. CHEN, E. KHURANA, A. SBONER, A. HARMANCI, J. ROZOWSKY, D. CLARKE, M. SNYDER AND M. GERSTEIN, *VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment*, Bioinformatics, 28(17):2267–2269, 2012.
- [39] K. J. KARZEWSKI, G. H. FERNALD, A. R. MARTIN, M. SNYDER, N. P. TATONETTI AND J. T. DUDLEY, *STORMSeq: An Open-Source, User-Friendly Pipeline for Processing Personal Genomics Data in the Cloud*, PlosOne, 9(1), (2014).
- [40] Z. WANG, Y. WANG, K. L. TAN, L. WONG AND D. AGRAWAL, *eCEO: an efficient Cloud Epistasis cOmputing model in genome-wide association study*, Bioinformatics, 27(8):1045–1051, 2011.
- [41] G. AGAPITO, M. CANNATARO, P. H. GUZZI, F. MAROZZO, D. TALIA AND P. TRUNFIO, *Cloud4SNP: Distributed Analysis of SNP Microarray Data on the Cloud*, In Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics (BCB'13).
- [42] P. H. GUZZI, G. AGAPITO, M. T. DI MARTINO, M. ARBITRIO, P. TAGLIAFERRI, P. TASSONE AND M. CANNATARO, *DMET-analyzer: automatic analysis of affymetrix DMET data*, BMC Bioinformatics, 13:258, 2012.
- [43] F. MAROZZO, D. TALIA AND P. TRUNFIO, *A cloud framework for big data analytics workflows on Azure*, In Proc. of the 2012 High Performance Computing Workshop, 2012.
- [44] F. MAROZZO, D. TALIA AND P. TRUNFIO, *Using clouds for scalable knowledge discovery applications*, In Euro-Par Workshops, pages 220227, Rhodes Island, Greece, 2012.
- [45] T. MUTH, J. PETERS, J. BLACKBURN, E. RAPP AND L. MARTENS, *ProteoCloud: A full-featured open source proteomics cloud computing pipeline*, Journal of Proteomics, 88:104–108, 2013.
- [46] E. AFGAN, D. BAKER, N. CORAOR, H. GOTO, I. M. PAUL, K. D. MAKOVA, A. NEKRUTENKO AND J. TAYLOR, *Harnessing cloud computing with Galaxy Cloud*, Nature Biotechnology, 29(11):972–974, 2011.
- [47] E. AFGAN, B. CHAPMAN, AND J. TAYLOR, *CloudMan as a platform for tool, data and analysis distribution*, BMC Bioinformatics 13(315), (2012).
- [48] L. JOURDREN, M. BERNARD, M. A. DILLIES AND S. LE CROM, *Eoulsan: a cloud computing-based framework facilitating high throughput sequencing analyses*, Bioinformatics, 11(28):1542–1543, 2012.
- [49] A. P. HEATH, M. GREENWAY, R. POWELL, J. SPRING, R. SUAREZ, D. HANLEY, C. BANDLAMUDI, M. E. MCNERNEY, K. P. WHITE AND R. L. GROSSMAN, *Bionimbus: a cloud for managing, analyzing and sharing large genomics datasets*, International Journal of American Medical Informatics Association, 2014, doi:10.1136/amiajnl-2013-002155.
- [50] S. V. ANGIUOLI, M. MATAKA, A. GUSSMAN, K. GALENS, M. VANGALA, R. RILEY, C. ARZE, J. R. WHITE, O. WHITE AND W. F. FRICKE, *CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing*, BMC Bioinformatics, 12(356), 2011.
- [51] *Data Intensive Academic Grid*, <http://diagcomputing.org/>.
- [52] *Magellan: Argonnes DOE Cloud Computing*, <http://magellan.alcf.anl.gov/>.
- [53] K. KRAMPIS, T. BOOTH, B. CHAPMAN, B. TIWARI, M. BICAK, D. FIELD AND K. E. NELSON, *Cloud BioLinux: pre-configured and on-demand bioinformatics computing for the genomics community*, Bioinformatics, 13(42), 2012.
- [54] J. P. EBEJER, S. FULLEA, G. M. MORRIS AND P. W. FINN, *The emerging role of cloud computing in molecular modelling*, Journal of Molecular Graphics and Modelling, 44:177–187, 2013.
- [55] A. WONG AND A. M. GOSCINSKI, *The Design and Implementation of the VMD Plugin for NAMD Simulations on the Amazon Cloud*, International Journal of Cloud Computing and Services Science (IJ-CLOSER), 1(4):155–171, 2012.
- [56] M. S. FRIEDRICH, P. EASTMAN, V. VAIDYANATHAN, M. HOUSTON, S. LEGRAND AND A. L. BEBERG, *Accelerating molecular dynamic simulation on graphics processing units*, Journal of Computational Chemistry, 30:864–872, 2009.
- [57] S. R. ELLINGSON AND J. BAUDRY, *High-throughput virtual molecular docking with AutoDockCloud*, Concurrency and Computation: Practice and Experience, 26(4):907–916, 2012.
- [58] G. FERNANDEZ-CARDENOSA, I. DE LA TORRE-DEZ, M. LPEZ-CORONADO, AND J. RODRIGUES, *Analysis of Cloud-Based Solutions on EHRs Systems in Different Scenarios*, Journal of Medical Systems, 36:3777–3782, 2012.
- [59] T. S. CHEN, C. H. LIU, T. L. CHEN, C. S. CHEN, J. G. BAU AND T. C. LIN, *Secure Dynamic Access Control Scheme of PHR in Cloud Computing*, Journal of Medical systems, 36(6):4005–4020, 2012.
- [60] C. DOUKAS, T. PLIAKAS, AND I. MAGLOGIANNIS, *Mobile healthcare information management utilizing Cloud Computing and Android OS*, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1037–1040, 2010.
- [61] S. BASU, A. H. KARP, J. LI, J. PRUYNE, J. ROLIA, S. SINGHAL, J. SUERMONDT AND R. SWAMINATHAN, *Fusion: Managing Healthcare records at cloud scale*, Computer, 45(11):42–49, 2012.
- [62] L. L. BOSCH-ANDERSEN, *Hospital uses cloud computing to improve patient care and reduce costs*,

- <http://www.microsoft.eu/2011/04/15/hospital-uses-cloud-computing-to-improve-patient-care-and-reduce-costs/>.
- [63] B. MENG, G. PRATX AND L. XING, *Ultrafast and scalable cone-beam CT reconstruction using MapReduce in a cloud computing environment*, Medical Physics, 38(12):6603–6609, 2011.
  - [64] M. S. AVILA-GARCIA, A. E. TREFETHEN, M. BRADY, F. GLEESON, F. AND D. GOODMAN, *Lowering the Barriers to Cancer Imaging*, Fourth IEEE International Conference on eScience (2008).
  - [65] L. A. BASTIO SILVA, C. COSTA AND J. L. OLIVEIRA, *DICOM relay over the cloud*, International Journal of Computer Assisted Radiology and Surgery, 8(3):323–33, 2013.
  - [66] R. P. PATEL, *Cloud computing and virtualization technology in radiology*, Clinical Radiology, 67:1095–1100, 2012.
  - [67] G. C. KAGADIS, C. KLOUKINAS, K. MOORE, J. PHILBIN, P. PAPADIMITROULAS, C. ALEXAKOS, P. G. NAGY, D. VISVIKIS AND W. R. HENDEE, *Cloud computing in medical imaging*, Medical Physics, 40, 2013, doi: 10.1118/1.4811272.
  - [68] J. HSIEH, A. LI AND C. YANG, *Mobile, Cloud, and Big Data Computing: Contributions, Challenges, and New Directions in Telecardiology*, International Journal of Environmental Research and Public Health, 10:6131–6153, 2013 doi:10.3390/ijerph10116131.
  - [69] H. XIA, I. ASIF AND X. ZHAO, *Cloud-ECG for real time ECG monitoring and analysis*, Computer Methods and Programs in Biomedicine, 110:253–259, 2013.
  - [70] G. FORTINO AND M. PATHAN, *Integration of Cloud computing and body sensor networks*, Future Generation Computer Systems 35:57–61, 2014.
  - [71] G. FORTINO, D. PARISI, V. PIRRONE AND G. DI FATTA, *BodyCloud: A SaaS approach for community Body Sensor Networks*, Future Generation Computer Systems, 35:62–79, 2014.
  - [72] G. FORTINO, R. GIANNANTONIO, R. GRAVINA, P. KURYLOSKI AND R. JAFARI, *Enabling effective programming and flexible management of efficient body sensor network applications*, IEEE Trans. Hum.Mach. Syst. 43(1):115133, 2013.
  - [73] J. D. PIETTE, M. O. MENDOZA-AVERALES, M. GANSER, M. MOHAMED, N. MARINE AND S. KRISHNAN, *A preliminary study of a cloud-computing model for chronic illness self-care support in an underdeveloped country*, American Journal of Preventive Medicine, 40(6), (2011).
  - [74] P. D. KAUR AND CHANA, I., *Cloud based intelligent system for delivering health care as a service*, Computer methods and programs in biomedicine, 113:346–359, 2014.
  - [75] T. W. KIM AND H. C. KIM, *A healthcare system as a service in the context of vital signs: Proposing a framework for realizing a model*, Computers and Mathematics with Application, 64:1324–1332, 2012.
  - [76] K. C. TSENG AND C. C. WU, *An Expert Fitness Diagnosis System Based on Elastic Cloud Computing*, The Scientific World Journal, 2014, <http://dx.doi.org/10.1155/2014/981207>.
  - [77] Y. LIN, K. L. PENG, J. CHEN, J. Y. TSAI, Y. C. TSENG, J. R. YANG AND M. H. CHEN, *Improvements in dental care using a new mobile app with cloud services*, Journal of the Formosan Medical Association, 2014, DOI: <http://dx.doi.org/10.1016/j.jfma.2014.02.009>.
  - [78] D. THILAKANATHAN, S. CHEN, S. NEPAL, R. CALVO AND L. ALEM, *A platform for secure monitoring and sharing of genome health data in the Cloud*, Future Generation Computer Systems, 35:102–113, 2014.
  - [79] M. JOHNSON, *Data hemorrhages in the health-care sector*, Financial Cryptography and Data Security, pp. 71–89, 2009.
  - [80] *Guidelines on security and privacy in public cloud computing*. National Institute of Standards and Technology (NIST), U.S. Department of Commerce. Special Publication, 800-144, <http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf>.
  - [81] S. KAMARA AND K. LAUTER, *Cryptographic cloud storage*, Financial Cryptography and Data Security, 2010, pp. 136–149.
  - [82] A. ABBAS AND S. U. KHAN, *A Review on the State-of-the-Art Privacy Preserving Approaches in the e-Health Clouds*, IEEE Journal of Biomedical and Health Informatics, 18(4):1431–1441, 2014.
  - [83] N. SULTAN, *Making use of cloud computing for healthcare provision: Opportunities and challenges*, International Journal of Information Management, 34:177–184, 2014.
  - [84] J. KNAUS, S. HIEKE, H. BINDER AND G. SCHWARZER, *Costs of cloud computing for a biometry department: a case study*, Methods Inf Med, 52(1):7279, 2013.
  - [85] D. D’AGOSTINO, A. CLEMATIS, A. QUARATI, D. CESINI, F. CHIAPPORI, L. MILANESI AND I. MERELLI, *Cloud Infrastructures for In Silico Drug Discovery: Economic and Practical Aspects*, Hindawi Publishing Corporation BioMed Research International, 2013, <http://dx.doi.org/10.1155/2013/138012>.

*Edited by:* Jesus Carretero

*Received:* September 16, 2014

*Accepted:* January 21, 2015