CONCEPTUAL FRAMEWORK OF BIO-CLOUD COMPUTING

[1] PROF. SWATI VITKAR, [2] PROF. VISHWANATH KALE

[1]Research scholar, JJT University, Jhunjhunu, Rajasthan
[2]Lecturer, S.I.E.S College, Nerul, Navi Mumbai, Maharashtra

swativitkar@gmail.com, kale_vishwanath@rediffmail.com

ABSTRACT: *Biology is experiencing an unprecedented growth of data and resources, what is making very difficult the access to local or conventional databases. New 'cloud computing' paradigm allows storing big data in remote servers and using web services to access and retrieve data efficiently. By doing so, researchers do not need to download, parse or integrate different sources since data are always up-to-date and can be retrieved by different client applications. BioCloud is a science-focused technology platform that will significantly advance the discovery of human therapeutics. It was designed to optimize the entire drug discovery process. BioCloud computing has emerged rapidly as an exciting new paradigm that offers a challenging model of computing and services. Leveraging cloud computing technology, bioinformatics tools can be made available as services to anyone anywhere and through any device. The use of large biodatasets, its highly demanding algorithms, and the hardware for sudden computational resources makes large-scale biodata analysis an attractive test-case for BioCloud computing.*

Keywords—Bio-cloud computing, SAAS, PAAS, IAAS

I: INTRODUCTION

BioCloud computing is becoming an adoptable technology for many of the organizations with its dynamic scalability and usage of virtualized resources as a service through the Internet. The bioinformatics cloud platform is designed to provide biological information and data analysis services for the biotechnology company, the current operational capacity of the platform is in the software as a service (SaaS), to the follow-up business development process, allowing users to participate in the construction of the cloud platform, including the development of a PaaS (Platform as a Service), IaaS (Infrastructure as a Service)-based platform functionality

This paper aims to explore the potential of "cloud computing for biology", and how it could be exploited in enhancing engagement among bioinformatics researchers and educators to better understand and improve their practice, in increasing the quality of their research outcomes.

II. CLOUD COMPUTING

Cloud computing is a computing model based on networks, especially based on the Internet, whose task is to ensure that users can simply use the computing resources on demand and pay money according to their usage by a metering pattern. Therefore, a new business model is being created where the services it provides are becoming computing resources [19].

Cloud computing consists of three layers.

1. Infrastructure as a service (IaaS)
2. Platform as a service (PaaS)
3. Software as a service (SaaS)

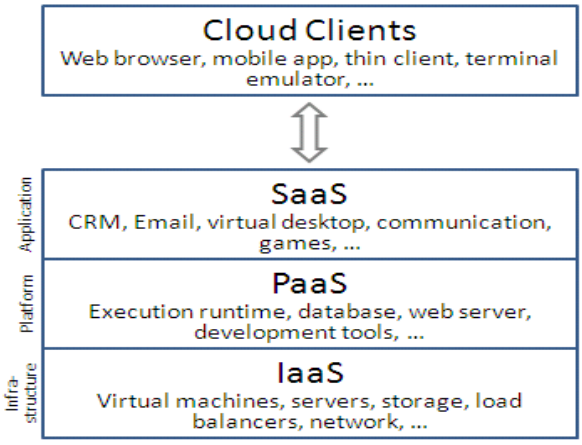

Fig. 1: Three Layers of Cloud Computing

Depending on the requirements, the customers can choose one or more services provided. The customers are renting or simply accessing the needed processing capacity from the data center using the client applications. The quality of the service becomes a crucial factor of the cloud computing success. Cloud computing is highly scalable and creates virtualized resources that can be made available to users. Users do not require any special knowledge about the concept of Cloud computing to connect their

computers to the server where applications have been installed and use them. Users can communicate through Internet with remote servers. These servers can exchange their computing slots themselves.
In Cloud computing, resources can be either externally owned (public Cloud – as provided by Google and Amazon) or internally owned (private Cloud). Public Clouds offer access to external users who are typically billed on a pay-as you-use basis. The private Cloud is built for the access within the enterprise where the users can utilize the facility without any charge.

A. Why biocloud Framework for Biologists?

Biocloud framework computing is an Internet-based biology process, using internet technology to design, implement, select, manage, support and extend biological processes, which will not replace traditional research and execution methods, but will greatly improve the efficiency of bio process implementations.

bioinformatics is widely used today on different organizational levels: for gene sequencing, Molecular programming, drug design etc. There are various bioinformatics solutions from open source to commercial. There are different entities involved in an bioframework design system: the researcher, Medicine companies and academicians.
The framework include the services like

- Cloud computing for biological databases
- Cloud computing for biological service
- Cloud computing for biological applications
- Cloud computing for drug design (i.e., modeling, virtual screening, docking)
- Cloud computing for genome analysis
- Cloud computing for computational biology techniques (i.e., sequence alignment, phylogenetic tree construction)
- Cloud computing for sequencing techniques and applications (i.e., resequencing, genome assembly, metagenomic)
- Cloud computing for RNA/protein structure
- Cloud computing for function prediction and classification
- Cloud computing for systems biology (i.e., domain-domain interaction, protein-protein interaction, pathways, network comparison, network alignment)
- Cloud computing for biological applications (i.e., viral research, infection disease)
- Cloud computing for biochip analysis (i.e., microarray, chip analysis)
- GPU technologies related to Biocloud

Usually, bioinformatics systems are developed as distributed applications, but this is not necessary so. The architecture of a distributed biocloud system includes software components, the client application, an application server and a database server and the necessary hardware components (client computer, communication infrastructure and servers).

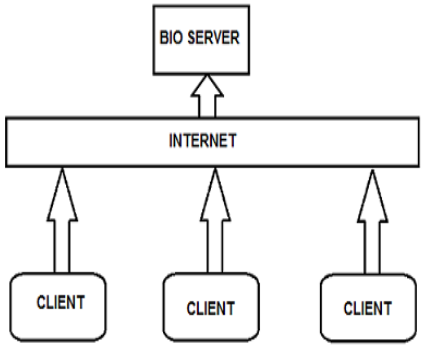

Fig. 2: Biocloud system

B. What bioCloud Can and Cannot do?

- Routine

– Small molecule conformation generation and energy profiling
– Visualizing crystal structures
– Binding site characterization
– Virtual screening to enrich databases for actives

- Cheminformatics, ligand-based, and structure-based

– Predict binding modes when receptor can be treated rigidly

- Difficult

– Separating highly from weakly active compounds
– Predicting side chain rearrangements and backbone relaxation

- Very Challenging

– Predicting binding free energies
– Predicting large scale protein movements
– Mapping free energy surfaces
– Understanding off-target effects

III. BIO-CLOUD COMPUTING ARCHITECTURE

The challenges that are faced by researcher are the ever changing environment. Economical issues and the requirement to reduce the operating cost to remain competitive. Cloud computing provides services that are easily configured, deployed, dynamically-scaled, and managed in virtualized environments. It provides the foundation and capabilities for the dynamic infrastructure through a consumption and delivery model for services in which the user sees only the service and has no need to know anything about the technology or implementation.

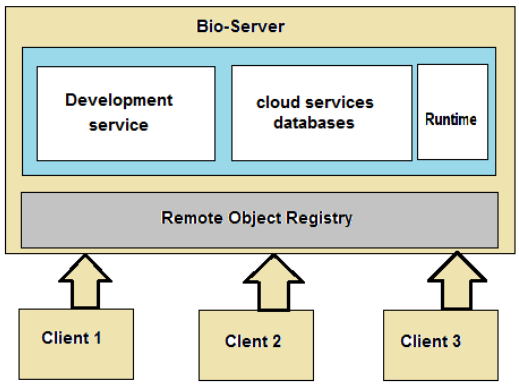

Fig. 3: BioCloud Computing Architectural Model

The key roles in a cloud Framework include the service requester, the service generator and the service provider. The cloud service consumer needs a secure anytime anywhere access to low cost services that are flexible and easy to use. The biggest hurdle to adoption of cloud has to do with consumers discomfort in the following areas: security of both service and the underlying data, service availability and reliability, service management to ensure service level agreements, ensuring control over access and policies, and the appropriate administration to facilitate flexible pricing structures. A service does not exist unless someone actually creates it. The cloud services creator needs tools and capabilities to offer differentiated services, offer incentives to ensure that consumers keep coming back to use the services, and the ability to change services on-demand to stay competitive and address threats. Finally the service provider actually runs the service that the service consumer wants and was designed and developed by the service creator.

The key capabilities of the reference architecture are defined in the Bio-server component. The lowest layer of the architecture defines the capabilities of the virtualized infrastructure. These capabilities facilitates virtualization of all biology: server, storage and network. These virtualization capabilities can handle all types of IT resources, e.g. both mainframe and distributed servers. The next layer provides an optimized remote object registry with capabilities for image deployment, integrated security, workload management, and high-availability. The optimized Remote object registry is used as the way to deliver services and information built according to well defined SOA and Information architectures. The central piece of the component is service management which provides the capabilities to manage a cloud service. These services include capabilities to handle user requests: managing the self service requests made by users, the lifecyle of images, and the provisioning of images based on the request. The capabilities also handle many of the qualities of service associated with delivering images, including availability, backup and restore, security and compliance, and performance management. To facilitate delivery through flexible models, the capabilities also support usage accounting and license management.

The reference architecture provides a comprehensive set of capabilities to ensure that cloud services can be built, deployed, accessed, delivered and managed. Each of these capabilities are supported by the appropriate standards, technologies and tools - all integrated to work together to deliver cloud computing. As shown in the figure, many of the components need management - from the virtualized resources through the workloads and into the images, usage and service requests. The virtualized resource management focuses on deploying cloud services on virtualized resources and managing cloud services, images, and resources. The workload scheduling and management focuses on defining workloads along with Qualities of Service (QoS) requirements, assigning resources to workloads, adding/removing resources based on workload needs, and monitoring workloads.

The service provider needs their IT resources integrated so their usage is optimized, the ability to add/remove resources on demand, a non-disruptive way to save money, and the means to charge for usage. There are three basic categories of cloud services: infrastructure services which provides access to a virtualized pool of resources, platform services that provide middleware or an application stack (hardware, operating system, software), and application services that provide access to a specific application or business process Where these services are deployed defines the type of cloud service. Private cloud services are hosted within the intranet while cloud services deployed in the internet are called public cloud services. It is critical to understand the workload characteristics to identify deployment strategies for a specific service.

The study reports that the cloud as a ubiquitous computing tool and a powerful platform can enable educators to practice new ideas. One of the most useful free “cloud computing” applications are the Google Apps for Education which is a free online suite of tools that includes Gmail for e-mail and Google Docs for documents, spreadsheets, and presentations. Using the cloud approach, everybody can work on the same document at the same time to make corrections as well as improve it dynamically in a collaborative manner.

In this paper a new paradigm is highlighted in bioinformatics area by introducing the cloud computing in order to increase the scalability, flexibility and availability of bio computing systems. The authors have evaluated the traditional networking model, with its advances and issues, and the possibility to move the e-learning system out of schools or enterprises, inside a cloud computing

infrastructure. The separation of entity roles and cost effectiveness can be considered important advantages. The research entity will be responsible for the deployment process, content management and delivery, and the vendor takes care of computational components, Resources, Data Storage, GUI's maintenance, development and management. The biocloud system can be scaled, both horizontally and vertically, and the bio medical clients is charged according to the number of used servers that depends on the number of services.

The analysis of cloud computing in higher education should be done in the both the views, ie., benefits and limitations.

Main Benefits and Limitations of Using Cloud Computing in biotechnology

Benefits

- Access to applications from anywhere
- Support for utilizing PDB's
- Software free or pay per use
- 24 hours access to infrastructure and content
- Opening to business environment and advanced research
- Protection of the environment by using green technologies
- Increased openness of students to new technologies
- Increasing functional capabilities

Limitations

- Not all applications run in cloud
- Risks related to data protection and security
- Organizational support
- Dissemination politics, intellectual property
- Security and protection of sensitive data
- Maturity of solutions
- Lack of confidence
- Standards adherence
- Offline usage with further synchronization opportunities
- Speed/lack of Internet can affect work methods

The main risks in cloud computing are security and data protection risks.

Issues regarding the security of cloud computing are:

- loss of governance
- lock-in
- isolation failure
- compliance risks
- management interface compromise
- data protection
- insecure or incomplete data deletion
- malicious insider
- protection of intellectual property and of the data in cloud

Security is a critical issue largely in public or shared environments, where the cloud provider needs to make sure that data privacy and compliance is guaranteed. Secure and efficient data exchange across the enterprise and clouds, as well as secure application connectivity are the major security concerns. Image management is important both in private and public clouds, as images are fast becoming the core object for deployment in data centers as a way to bypass installation problems. In this context, organizations need a way to organize, secure, manage and deploy images to the various virtualized platforms in a scalable manner. Once deployed, organizations need a way to manage the virtual images, which includes monitoring, updating, tracking, change management and auditing

In case of biotech institutions sensitive data must be paid special attention. In cloud there are many solutions for ensuring security and protection of data like mask the data, firewalls and encryption and key management. To choose solution for data protection and security the risk and cost of implementation must also be considered. To protect data against unauthorized access in the cloud environment data encryption can be used. The data can be encrypted before transferring it in the cloud.

IV. CONCLUSION

As a conclusion of Cloud Computing in biology, we may say that the payment per use model and the management policies of risks and security represent positive factors in taking the decision of using Cloud Computing. Moreover, Cloud Computing adds value with small capital expenses, assuring at the same time the protection of the environment.

By including the cloud services, biotech organizations achieve a substantially decreasing of expenses with software licensing and at the same time to reduce the infrastructure by 75% with full working schedule.

The use of Cloud Computing becomes a necessity and not an option for many biotech companies. This aspect is due to a various factors such as costs increase, the pressure of income increase, experimental success, institutional performance and competition.

REFERENCES

[1] N. Atkinson, ―NASA creates a new NEBULA: Cloud computing project,‖ Universe Today, June 4, 2009.

[2] Paul Pocatilu, Felician Alecu, Marius Vetrici, Measuring the Efficiency of Cloud Computing for E-learning Systems, WSEAS TRANSACTIONS on COMPUTERS

[3] IBM Academy of Technology Thought Leadership White Paper, October 2010, Cloud computing insights from 110 implementation projects IBM Academy of Technology Survey

[3] DeCoufle B., The impact of cloud computing in schools, The Datacenter Journal, http://datacenterjournal.com/content/view/3032/40/, July 2009
[5] Boja C., Batagan L., Software Characteristics of M-Learning Applications, Proc. of. 10th WSEAS International Conference on Mathematics and Computers in Business and Economics (MCBE'09), Prague, Czech Republic, March 23-25, 2009, ISSN: 1790-5109, ISBN: 978-960-474-063-5, pp. 88-93;
[6] Marinela Mircea and Anca Ioana Andreescu, Using Cloud Computing in Higher Education: A Strategy to Improve Agility in the Current Financial Crisis, IBIMA Publishing Communications of the IBIMA http://www.ibimapublishing.com/journals/CIBIMA/cibima.html Vol.2011, Article ID 875547, 15 pages DOI: 10.5171/2011.875547
[7] T. Jian, F. Lijian, G. Tao, "Cloud computing-based Design of Network Teaching System", Journal of TaiYuan Urban Vocational college, Mar. 2010, pp.159-160.
[8] Y. Zhongze, "The basic principles of cloud computing and its impact on education", Satellite TV and Broadband Multimedia, 2010.6, pp.67-70.

[9] Z. Chengyun, "Cloud Security: The security risks of cloud computing, models and strategies", Programmer, May.2010, pp.71-73.
[10] B. Hayes, "Cloud computing," Comm. Acm, vol. 51, no. 7, pp. 9– 11, 2008.
[11] R. Buyya, C.S. Yeo & S.Venugopal, "Market-oriented Cloud computing: Vision, hype, and reality of delivering IT services as computing utilities," 10th Ieee Int. Conf. High Performance Comput. Comm., p. 5–13, 2009.
[12] M. Lijun, W.K. Chan & T.H. Tse, "A tale of Clouds: Paradigm comparisons and some thoughts on research issues," Ieee Asia-pasific Services Comput. Conf., Apscca08, pp. 464–469, 2008.
[13] K. Praveena& T. Betsy, "Application of Cloud Computing in Academia," Iup J. Syst. Management, vol. 7, no. 3, pp. 50–54, 2009.
[14] K.A. Delic & J.A. Riley, "Enterprise Knowledge Clouds," Next Generation Km Syst. Int. Conf. Inform., Process, Knowledge Management, Cancun, Mexico, pp. 49–53, 2009.
[15] J. A. Méndez and E. J. González, "Implementing Motivational Features in Reactive Blended Learning: Application to an Introductory Control Engineering Course", IEEE Transactions on Education, Volume: PP, Issue: 99, 2011.
[16] D. Chandran and S. Kempegowda, „Hybrid bioserver Platform based on Cloud Architecture Model: A Proposal", Proc. International Conference on Signal and Image Processing (ICSIP), pages 534-537, 2010.
[17] L. Huanying, "Value and understanding for cloud computing based on middleware", Programmer, 2010.05. pp.68,69