Current Trends in Biotechnology and Pharmacy
Vol. 11 (3) 294-299 July 2017, ISSN 0973-8916 (Print), 2230-7303 (Online)

Cloud Computing and its Need in Life Science's R & D

L. N. Chavali[1], N. Lakshmi Prashanthi[2], K. Sujatha[3] and P. B. Kavi Kishor[1]*
[1]Department of Genetics, Osmania University, Hyderabad - 500 007, India
[2]GITAM Institute of Management, GITAM University, Rushikonda, Visakhapatnam 530 045, Andhra Pradesh, India
[3]Department of Chemistry, KTR Women's College, Gudivada- 521 301, Andhra Pradesh, India
*For correspondence - pbkavi@yahoo.com

Abstract

The cloud computing is driven by factors such as aging of the current information technology (IT) infrastructure and changes in IT landscape. Virtualization and cloud computing have transformed the way the IT services are delivered to customers at a lower cost. The research and development (R&D) in life sciences is on cusp of change. The new opportunities and challenges must focus on renewing the existing systems and processes for efficacy and adopting advancements in new technologies to gain value. This paper briefly outlines various cloud computing trends in life sciences.

Introduction

Cloud computing means services carried over internet from anywhere, any place, anybody and any device. A new era is set for an integrated development and the convergence of various technologies such as social media, advanced analytics, digital media, and Big Data with the cloud and explosion of digital data availability due to this convergence play a key role in R&D in life sciences (1, 2). The cloud computing solutions, especially software-as-a-service (SaaS), are reshaping the service models in life sciences. The disruptive technology is considered as the game-changer for life sciences (1, 2). Ultimately, several companies using 'SaaS-first' approach for R&D initiatives and eventually, an "all-in" cloud SaaS scenario will open up promising opportunities related to innovations in drug developments and personalized medicines (1, 2).

Cloud provides three types of services to its users *viz.*

a) Software as a Service (SaaS) – enables users to run applications online through cloud computing techniques Ex: Google mail, Google calendar API, salelsforce.com etc.

b) Platform as a Service (PaaS) – enables the users to build their own cloud applications using supplier-specific tools and languages through cloud computing techniques. Ex:MSSQL, Appche, Linux etc.

c) Infrastructure as a Service (IaaS) – enables the users to run any application on supplier's cloud hardware. Ex: Amazon EC2, GoGrid etc.

Shared resources environment in cloud offers faster computation, significant economies of scale to cut cost and improve operational agility. Cloud computing provides opportunities to move away from capital investments to operational costs (3). Cloud solutions are generally architected for multi-tenancy, scalability and reliability. Instead of buying the big servers, computers, and software packages, cloud computing, like electricity, becomes an operational cost and any analysis can be performed in a few weeks for a few thousand dollars and also moving away from desktop installed packages for services (3). As cloud offers inter organizational or departmental collaboration, migrating to cloud can be a strategic differentiator especially for pharmaceutical companies (3).

The innovation through cloud in life sciences is due to

a. Data accessibility
b. Low cost
c. Infrastructure provisioning
d. More computing power
e. Enhanced collaboration

R&D informatics, social media analytics and digital health technology are creating paradigm shift in life sciences. Cloud computing offers the use of data encryption, password protection, secure data transfer, processes' audits, and the implementation of respective policies against data breeches and malicious use. There is concern and flawed assumption about data privacy and security in public clouds but there is a need to focus more on developing enterprise cloud strategy which includes security in terms of security -aware application design, application self-protection besides traditional measures than worrying about it. Many standards have evolved over a period of time to mitigate the risks. The standard bodies like National Institute of Standards and Technology (NIST), Open Cloud Computing Interface (OCCI), Open Cloud Consortium etc., are working on issues such as data security and data integration (3).

Gartner Inc.,(USA) predicted four outcomes for postmodern business applications, using the hybrid reality, on-premises monolith, outsourced everything, and flip (HOOF) model (Fig. 1). Hybrid is complementing the existing on-premises business applications with cloud-delivered business applications and Flip is "Flipping" the majority of on-premises business applications to the cloud.

Cloud computing trends

The cloud computing in Life Sciences R&D is primarily driven by a surge in multistate clinical trials, wide prevalence of chronic diseases in major regions, and researches in gene therapy and the innovations in drug making such as those of personalized medicines (2, 3). The data explosion due to the rise of genomics has accelerated the adoption of cloud computing in R&D. In addition to the genomic data explosion, the harnessing of real world data from wearables are contributing to the massive data processing

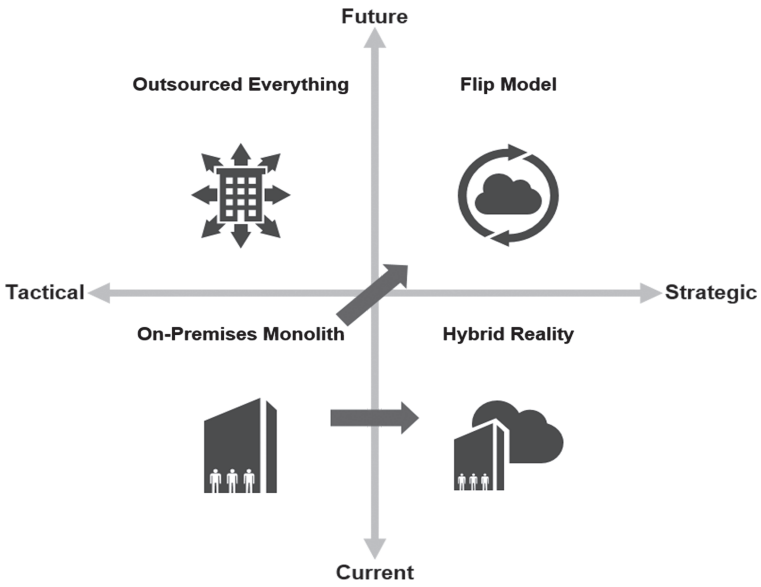

Fig. 1. Gartner's HOOF Model-Options for cloud deployment of business applications
Source: Gartner Inc (August 2016) (with permission from Gartner Inc)

needs which make cloud computing critical. Therefore, adoption of cloud in life sciences enables new ways of working by capturing new opportunities and by transforming life science industry for improving the quality of life (2, 3).

Cloud computing in bioinformatics

It was thought in the beginning that by generating huge amounts of data, it is possible to process and compute the DNA variants by comparing them to the reference genome but storing terra bytes of data and using it for analysis at a later date for updating the variant profile is a huge challenge. Data interpretation and integration is the key in innovation (4-7). Modern biology faces challenges in the field of data storing, management and analysis. Sequencing and cataloguing genetic information has increased many folds (4-7). Various medical research institutes are continuously targeting on sequencing of millions of genomes for the understanding of biological pathways and genomic variations to predict the cause of the disease and developing genome-based therapeutics. The explosion of biological data (the scale of the data exceeds a single machine) has made it more expensive to store, process and analyze compared to its generation. This has stimulated the use of cloud to avoid large capital infrastructure and maintenance costs. The chief advantage to a researcher in biology using cloud computing is the capacity to scale the analysis up and complete it in short period of time. Due to explosion of data, the logic of analysis will be taken near to data and not the other way round. DNA sequencing generates massive amount of data, and its interdisciplinary nature employs cloud computing and big data technologies in life sciences. Next generation cloud based computational biology has the potential to revolutionize life sciences (4, 5,6, 7). Genomics research can lead to a greater height by leveraging the advantages that the current technologies offer. The solutions in cloud computing are expanding rapidly day by day in computational biology (4, 5,6, 7). Cloud computing is currently utilized for:

a. bioinformatics workflows
b. comparative genomic studies
c. gene set analysis for biomarkers
d. identification of epistatic interactions between single-nucleotide polymorphisms
e. microbial/animal/plant sequence analysis
f. multiple sequence alignment algorithms
g. protein annotations and analysis
h. and systems biology

Despite the wide range of different cloud computing platforms available, most of the existing works in computational biology are focused on Amazon Web Services (AWS) as provided by Amazon, Amazon Web Services (AWS). Amazon provides a centralized cloud of public data sets (e.g. archives of GenBank, Ensembl databases, etc.) of biology; chemistry etc., as services and it is the biggest cloud provider for big data processing. Beijing Genomics Institute (BGI), the world's largest genomics organization, introduced its latest-generation cloud-based Software as a Service (SaaS) solution, EasyGenomics™. EasyGenomics integrates various popular next generation sequencing (NGS) analysis workflows including whole genome re-sequencing, exome re-sequencing, RNA-Seq, small RNA, *de novo* assembly, among others (4, 5). EasyGenomics can be accessed via www.easy genomics.com. Galaxy is an open source next generation sequencing platform. Similarly, NextBio is a search tool to mine all the public databases available. DNAnexus is another cloud-based system for NGS data storage and analysis (4, 5). Sage Bionetworks has developed an open source synapse platform for managing and sharing data and computational workflows. Synapse is a platform for supporting scientific collaborations centered on shared biomedical data sets. The primary goal is to make biomedical research more transparent, reproducible, and accessible to a broader audience of scientists.

Cloud computing in pharmaceutical industry

The Research and Development process in pharmaceutical industry is complex and there are many specialized applications of different brands supporting R&D. These applications, which are in silos, generate huge amount of data which

require maintenance due to Federal Drug Agency (FDA) regulations. The companies are looking for ways to reduce the IT budget and hence, are front runners of cloud adoption (2, 3). There is a growing trend to simplify platforms and solutions in areas such as Enterprise resource planning (ERP) environment and non-ERP systems. Typical pharmaceutical applications in bioinformatics such as DNA and protein sequence analysis, machine learning, data mining and applications in chemical informatics such as virtual screening, Quantitative Structure Activity Relations (QSAR) etc., are available as SaaS in cloud. Molecular modeling and simulation, Bio Assay, Basic Local Alignment Search Tool (BLAST) etc., along with other enterprise applications such asERP, Customer Relationship Management (CRM) are hosted in cloud. Besides, a switch to cloud-based solutions in clinical IT has been taking place over the last two-to-four years. Players like Johnson and Johnson (J&J), Pfizer companies have already adopted cloud and started porting their data to cloud.

Cloud computing in agriculture (Mobile agriculture)

Low yield per hectare, accelerated agricultural growth, sustainable water management, combating dry land salinity, poor access to proper infrastructure, credit and modern technology etc., have created new challenges in agricultural farming. In the beginning, Information and Communication Technology (ICT) was adopted to reduce the digital divide in rural populations. But, providing service on demand has pushed the industry and scientists to share knowledge and collaborate which has resulted in adoption of cloud. Spread of wireless networks, low cost mobile phones and high reach of Wi-Fi networks at rural areas made it easy to adopt to cloud. Mobile agriculture (M-agriculture) is playing a crucial role in information gathering and analysis. United States Department of Agriculture (USDA) has transitioned 1,20,000 federal workers from on-premise messaging and collaboration to Microsoft's cloud computing solution with a vision to consolidate disparate messaging environments

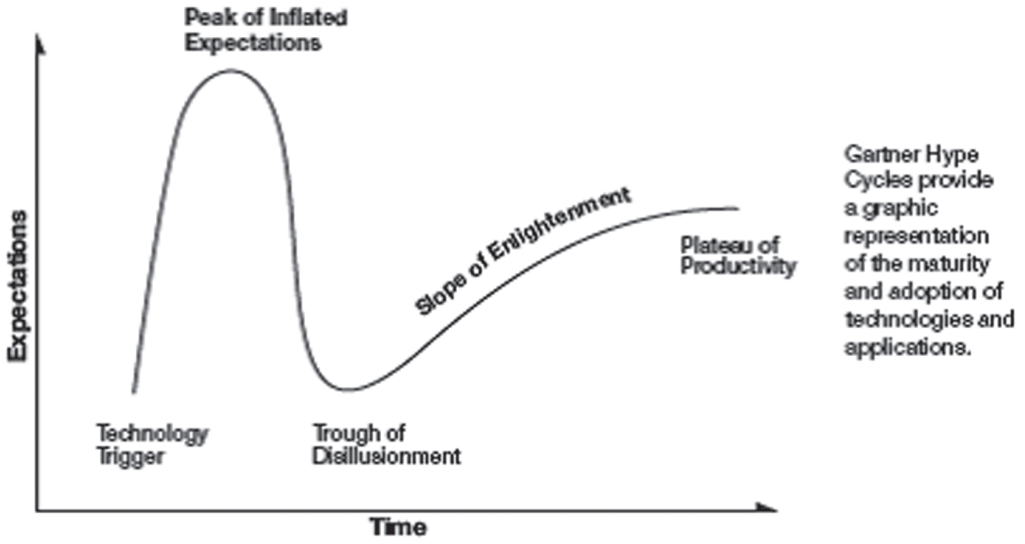

Fig. 2. Hype Cycle– Source: Gartner Inc (August 2016) (with permission from Gartner Inc).

onto a single, unified platform, which will reduce the costs, boost workforce productivity and improve communications and collaboration. International System for Agricultural Science and Technology (AGRIS) is a global public domain database with more than 4 million structured bibliographical records on agricultural science and technology (7). This database is maintained by FAO, and its content is provided by more than 150 participating institutions from 65 countries. Hopefully, it would be more useful to the farming community in the years to come.

The five phases of the life cycle are explained below:

1. Technology trigger - The hype is created by new technology innovation or breakthrough.
2. Peak of inflated expectations - The product generated during this phase may be specialized, costly and have high expectations. More players jump in this phase to understand the fitment and behavior of the new technology in their respective business.
3. Trough of disillusionment - This phase consists of negative publicity and pushing technology beyond its limits.
4. Slope of enlightenment -The technology begins its climb toward the early stages of maturity as second- and third-generation products are launched, and methodologies and tools are added to ease the development process.
5. Plateau of productivity - The plateau of productivity represents the beginning of mainstream adoption, when the business benefits of the technology are positively measured.

The competition that builds up at peak of inflated expectations and the trough of disillusionment may slowly lead to consolidation. Software as a Service (SaaS) is getting more matured and entering into plateau of productivity due to availability of cloud migration tools, cloud testing tools, application PaaSetc (8). The cloud computing and SaaS are the key differentiators for optimizing cost as well as promoting innovation in research (8).

Outlook

Figure 2 describes the hype cycle of cloud computing for the next decade. The emergence of SaaS as a front line model is evident from this figure. AI and machine-learning capabilities are seeping into virtually every technology, and represent a major battleground for technology providers over the next ten years. The next generation may be driven by blending digital and physical worlds with artificial intelligence (AI) and machine learning. It can be assumed that every application or service will incorporate AI in some way over the next 10 years as it can be seen now that some of service providers are increasingly using AI and machine learning in their applications. In life sciences, AI and machine learning are widely used now in advanced analytics. DevOps tools help to overcome the barriers between security and application teams. All future application/service design will embrace DevSecOpsmodel. The future SLA must focus on (a) dynamic storage provisioning in terms of capacity, bandwidth and throughput and coexistence of private and public clouds (b) CPU/ memory virtualization and (c) workflow/service SLAs. CSCC (Cloud Standards Customer Council) is actively working on building the artifacts in leveraging of cloud by various workload types including Big Data, social and mobile. To scale trillions of services in cloud, robust standards in terms of global interoperability and massive scaling are required which may be evolved and defined by standard bodies like NIST, DMTF, SNIA, OGF, IEEE, etc.

References

1. Kavi Kishor, P.B., Rajib Bandopadhyay and Prashant S. Editors (2014). Agricultural Bioinformatics. Springer Verlag, New Delhi, Heidelberg, New York, Dordrecht, London. ISBN: 978-81-322-1879-1. ISBN: 978-81-322-1880-7 (eBook).189-213.
2. Chang, V., Walters, R.J. and Wills, G.B. (2015). Cloud computing and frameworks for

organizational cloud adoption. In: *Delivery and Adoption of Cloud Computing Services in Contemporary Organizations*.IGI Global, Ed. By Chang, Victor, Walters, Robert John and Wills, Gary (2015) ePrint ID: 366811.

3. OECD Council Meet, Paris (2016). Enabling the next production revolution: the future of manufacturing and services - Interim report. pp14-19.
4. Shanahan, H.P., Owen, A.M. and Harrison, A.P. (2014). Bioinformatics on the cloud computing platform azure. PLoS ONE 9(7): e102642. doi:10.1371/journal.pone.0102642.
5. Krampis, K. and Wultsch, C. (2015). A review of cloud computing: bioinformatics solutions for next-gen sequencing data analysis and research. Methods Next-Generation Seq. 2: 23-34.
6. O'Driscoll, A., Daugelaite J. and Sleator, R.D. (2013). 'Big data', Hadoop and cloud computing in genomics. J Biomed Inform 46: 774-781.
7. Chavali, L.N., Hima Kumari, P., Punitha, D., Prashanthi, N.L., Prashanth, S. and Kavi Kishor, P. B. (2013). Database Methods in Software-as-a-Service (SaaS) in Agricultural Services. Proceedings of Andhra Pradesh Akademi of Sciences 15: 93-100.
8. Akhmatovich, T.K., Ugli, I.S.Z. and Odiljon, Z. (2016). Analyze threats in cloud computing. Journal of Electrical and Electronic Engineering 4: 145-149.