International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 5 Issue: 5
ISSN: 2321-8169
1256 – 1259

Exploit Cloud Computing for Next Generation Sequencing Analysis

Mamta Arora[1],
[1]Associate professor & HOD Biotechnology, ASBASJSM College, Bela Rupnagar (Punjab)
mamtaarora.2007@rediffmail.com

Parvinder Kaur[2]
[2]Assistant Professor
Biotechnology, ASBASJSM College, Bela Rupnagar (Punjab)
harrysaini88@gmail.com

Abstract: Over the last decade more data has been produced than in the entire human civilization. Advances in the field of sequencing techniques have accelerated the process leading to huge sequence database. This presents immediate challenge not only in database management but computational analysis also. With the advent of next generation sequencing data bases, biologist are highlighting the immediate need for computer power, storage and software's to analyse this data.This can be addressed by moving from in house technology to innovative analysis technology i.e. Cloud computing. Though this technology is in bud stage in NGS but it provide a promising future for next generation sequencing. Data management, transfer, storage, security, computing power are all important components to success in using cloud computing for next generation sequence analysis. This paper is written to bring awareness to scientific community regarding promoting NGS analysis by using cloud computing.

Keywords; *next generation sequencing (NGS), softwares, Cloud computing, Bioinformatics*

*****

I INTRODUCTION

Next generation sequencing (NGS) is an advanced technology that is playing a landmark role in molecular genetic testing of humans. NGS methods carried out DNA sequencing in bulk and produces results in short period of time at relatively less cost. There are various NGS methods including Illumina, Pacific Biosciences, Roche 454 and Nanopore technologies [1]. DNA sequencing generates large amount of data in the form of small DNA sequences called as reads. Here comes the need of Bioinformatics, a discipline of Science that uses latest computational tools to analyse and interpret large amount of biological data.In NGS, bioinformatics tools are developed to convert signals into data, data into interpretable information, and information into actionable knowledge. This process can be categorized into primary, secondary and tertiary analyses (Fig.1)

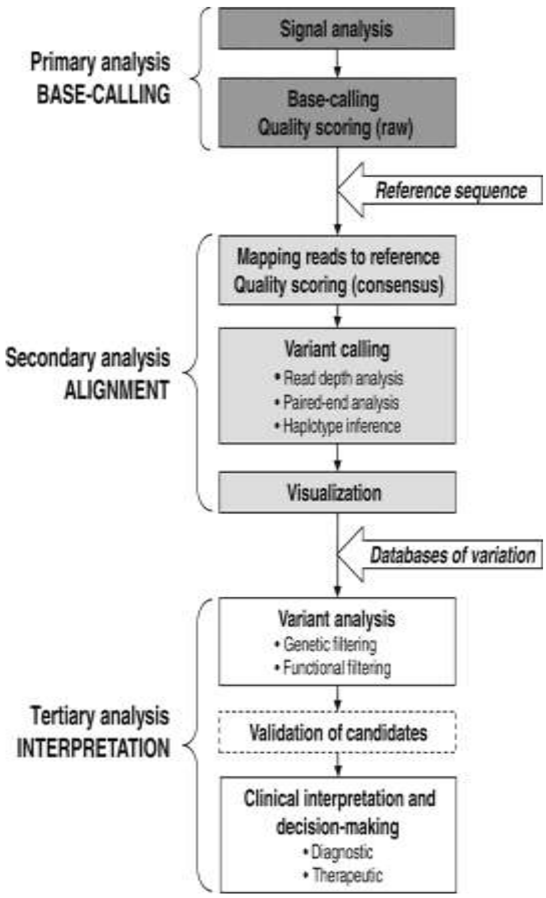


Figure1 . Outline of primary, secondary and tertiary analyses in NGS [2].

1. Primary analyses: It involves the conversion of raw data into short sequences of DNA on the basis of changes in electrical current or light intensity [3].
2. Secondary analyses: It involves the comparison of individual DNA sequence with reference sequence and determining any variation from reference sequence [4].
3. Tertiary analyses: It includes the analyses of variants, their origin and functional impacts [5]. Commonly used software in NGS are enlisted in table1.

Table1. Commonly used softwares for NGS [3, 4,6-9]

<table>
<tr><th>Functions</th><th>Software</th><th>Cloud Computing Services</th></tr>
<tr><td>Alignners</td><td>SAMtools,
EagleView
,MaqView,
Tablet,
MapView,
IGV</td><td rowspan="7">Commercial source

Bina technologies
DNA nexus
Easy genomics
Gene sifter
Life Sciences
Seven bridges

Open source
Cloud Aligner

Cloud Burst
clovR
Croosshow
FX
Galaxy
SIMPLEX
TREAT
VAT</td></tr>
<tr><td>Analysis of evolutionary conservation</td><td>BLAST,
BLAT,
phyloP,
PHAST</td></tr>
<tr><td>Analysis of Single nucleotide polymorphism</td><td>Bowtie,
BWA,
GATK,
MAQ,
MOSAIK,
SAMtools,
SSAHA2,
SOAP2</td></tr>
<tr><td>Genome assembly (Cancer)</td><td>Mutect
JoinSNVMix2
SNVmix,
SomaticSniper,
TIGRA,
VarScan</td></tr>
<tr><td>Amino acid substitution effects</td><td>Align GVGD
nsSNPA,
MAPP,
PANTHER,
PhDSNP,
PMUT,
PoyPhen-2,
SIFT,
SNPs3D</td></tr>
<tr><td>Structural variants</td><td>BreakDancer,
CREST
DELLY
Dindel,
Pindel,
PEMer,
VariationHunter</td></tr>
<tr><td>Visualization of aligned data</td><td>Pairoscope</td></tr>
</table>

II. WHAT IS CLOUD COMPUTING?

It is a model[9]that provides the client on demand services through internet. These include networks, servers, storage, applications and services. In nut shell it has 5 characteristics.

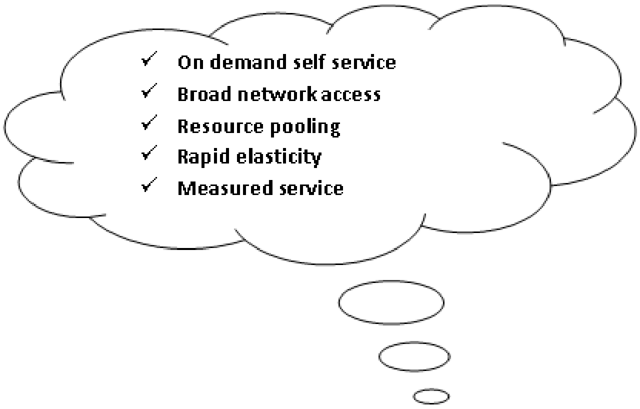


Figure 2 : characteristics of Cloud Computing

A. Types of Cloud: There are 3 types of clouds private, public and hybrid clouds. Priavte clouds are maintained within an organization and owned by an enterprise solely for its own use. in public clouds, Cloud provider rents services on demand. Hybrid cloud are combination of public and private clouds.

B. Services provided by cloud computing: services can be categorised under 3 categories[10]:
Software as a services (SaaS): Software is provided on demand to client.
Platform as a service (PaaS) working platform is provided.
Infrastructure as a service (IaaS)this provides computing capabilities and basic storage over the network.
Cloud computing services for NGS: These services are either commercial or open source tool as shown in table 1.

III. ADVANTAGES OF CLOUD COMPUTING

Advantages [11,12]can be summarised as in Figure 3.

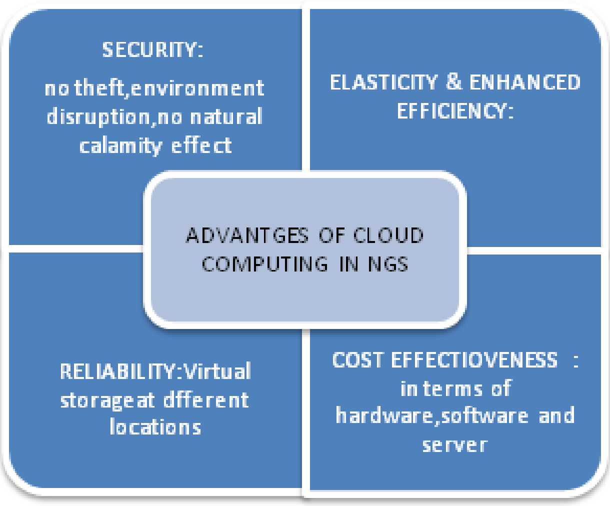
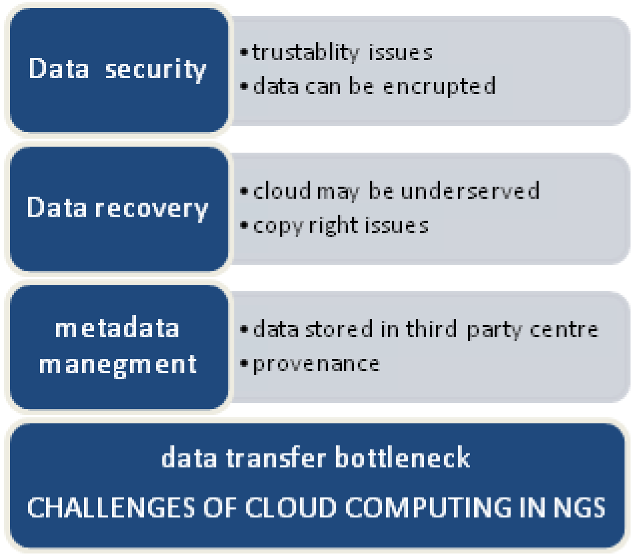


Figure 3: Advantages of Cloud Computing in NGS

IV. CHALLENGES

Although a lot of advantages are there, still challenges have to be overcome to get worthwhile results. therefore scientists must exploit this technology for the betterment of community.



V. CONCLUSION

Despite the challenges existing in cloud computing. efforts must be devoted to exploit cloud computing to whole community to accelerate next generation sequencing data analysis for diagnosis, prognosis, drug discovery and drug development.

REFERENCES

[1] S. Goodwin, J. D. McPherson, and W. R. McCombie, "Coming of age: ten years of next-generation sequencing technologies," Nat. Rev. Genet., vol. 17, no. 6, pp. 333–351, June 2016.
[2] S. Moorthie, A. Hall, and C.F. Wright," Informatics and clinical genome sequencing: opening the black box, Genet. Med., vol. 15, pp. 165-171, September 2013.
[3] S. Moorthie, C. Mattocks, and C. Wright," Review of massively parallel DNA sequencing technologies,"The HUGO J., vol. 5, pp. 1-12, December 2011.
[4] P. Flicek, E. Burney," Sense from sequence reads: methods for alignment and assembly," Nat. Methods, vol. 6, pp. 6-12, November 2009.
[5] G. Kuhlenbaumer, J. Hullmann, and S. Appenzeller," Novel genomic techniques open new avenues in the analysis of monogenic disorders," Hum.Mutate. vol. 32, pp 144-151, February 2011.
[6] S. Boa, R. Jang, W. Kwan, B. Wang, X. Ma, and Y.Q. Song,"Evaluation of next-generation sequencing software in mapping and assembly,"J. Hum. Genet., vol. 56, pp.406–414, June 2011.
[7] A. Magi, M. Benelli, A. Gozzini, F. Girolami, F. Torricelli, and M. Brandi," Bioinformatics for next generation

sequencing data," Genes, vol. 1, pp. 294–307, September 2010.

[8] M. Pop, and S. L. Salzberg," Bioinformatics challenges of new sequencing technology,"Trends Genet., vol. 34, pp 142-149, March 2008.

[9] T. Kwon, W. G. Yoo, W.-J. Lee, W. Kim, and D.-W. Kim, "Next-generation sequencing data analysis on cloud computing," *Genes Genomics*, vol. 37, no. 6, pp. 489–501, 2015.

[10] X. Guo, N. Yu, B. Li, and Y. Pan, "Cloud Computing for Next-Generation Sequencing Data Analysis," *Comput. Methods Next Gener. Seq. Data Anal.*, no. March, pp. 1–24, 2016.

[11] C. Mawere, K. Zvarevashe, and T. Sengudzwa, "Cloud - Based Big Data Analytics in Bioinformatics : A Review," pp. 217–230.

[12] R. S. Thakur, R. Bandopadhyay, B. Chaudhary, and S. Chatterjee, "Now and next-generation sequencing techniques: Future of sequence analysis using cloud computing," *Front. Genet.*, vol. 3, no. DEC, pp. 1–8, 2012.