AN ANALYSIS OF E HADOOP/MAPREDUCE/H BASE FRAMEWORK AND ITS CURRENT APPLICATIONS IN BIOINFORMATICS

Ramesh Chandra Tripathi*

*Professor,
Department of Computer Engineering, Teerthanker Mahaveer University,
Moradabad, Uttar Pradesh, INDIA
Email id: tripathi.computers@tmu.ac.in

DOI: 10.5958/2249-7137.2021.02501.5

ABSTRACT

High-performance computing (HPC) has become more essential in bioinformatics data processing as a result of new computational difficulties. Work is usually distributed over a cluster of computers that connect to a shared file system housed on a storage area network. The Message Passing Interface (MPI) and, more recently, Hadoop's MapReduce API have been used to achieve work parallelization. Cloud computing is another computer architecture/service model that is currently being investigated. In a nutshell, cloud computing is HPC with a web interface plus the flexibility to scale up and down quickly for on-demand usage. Remote clients upload potentially large data sets for analysis in the Hadoop framework or other parallelized environments running in the data center, with the server side deployed in data centers working on clusters. The present use of Hadoop, a toplevel Apache Software Foundation project, and related open source software projects in the bioinformatics field is discussed. The principles underlying Hadoop and the HBase project are explained, as well as the existing bioinformatics software that uses Hadoop. The emphasis is on next-generation sequencing, which is now the most popular application area.

KEYWORDS: *API, Hadoop, H Base, Map Reduce, Pig.*

REFERENCES:

1. Diaconita V, Bologa AR, Bologa R Hadoop oriented smart cities architecture. Sensors (Switzerland), 2018, doi: 10.3390/s18041181.
2. O'Driscoll A, Daugelaite J, Sleator RD. Big data', Hadoop and cloud computing in genomics. Journal of Biomedical Informatics. 2013;46(5):774–781. doi: 10.1016/j.jbi.2013.07.001.
3. Hodge VJ, O'Keefe S, Austin J. Hadoop neural network for parallel and distributed feature selection. Neural Networks, 2016 Jun;78:24-35. doi: 10.1016/j.neunet.2015.08.011.
4. Taylor RC. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. BMC Bioinformatics, 2010;11(Suppl 12):S1. doi: 10.1186/1471-2105-11-S12-S1.
5. White T. Hadoop: The definitive guide 4th Edition. Online, 2012, doi: citeulike-article-

id:4882841.

6. Sirisha N, Kiran KVD. Authorization of data in Hadoop using Apache Sentry. Int. J. Eng. Technol., 2018;7(3.6): 234-236. doi: 10.14419/ijet.v7i3.6.14978.

7. Niemenmaa M, Kallio A, Schumacher A, Klemelä P, Korpelainen E, Heljanko K. Hadoop-BAM: Directly manipulating next generation sequencing data in the cloud. Bioinformatics, 2012 Mar 15;28(6):876-7. doi: 10.1093/bioinformatics/bts054.

8. Polato I, Ré R, Goldman A, Kon F. A comprehensive view of Hadoop research - A systematic literature review. Journal of Network and Computer Applications. 2014;46:1–25. doi: 10.1016/j.jnca.2014.07.022.