Abstract

Apache Hadoop is an emerging technology that is widely used in the data intensive applications like Big Data Analysis. This technology is currently used in the searching applications of Google, Yahoo, and Amazon. But recently it is found that Hadoop is widely being used in the data analysis of genome sequences of Biosciences. Bioinformatics is the new field of science where the computer technology is used in studying the molecular biology. For example, Hadoop technology can be used in finding a particular genome sequence from a huge datasets of Genes in Genetics.

References

2. “HOG: Distributed Hadoop MapReduce on the Grid” Chen He, Derek Weitzel, David
Swanson, Ying Lu Computer Science and Engineering University of Nebraska– Lincoln, Email: fche, dweitzel, dswanson, ylug@cse.unl.edu.

3. “CloudBLAST: Combining MapReduce and Virtualization on Distributed Resources for Bioinformatics Applications”, Andréa Matsunaga, Maurício Tsugawa and José Fortes Advanced Computing and Information Systems Laboratory Department of Electrical and Computer Engineering, University of Florida PO Box 116200, Gainesville, FL, 32611-6200, USA{ammatsun, tsugawa, fortes}@ufl.edu. Fourth IEEE International Conference on eScience.

4. “CloudBurst: highly sensitive read mapping with MapReduce”, Michael C. Schatz* Center for Bioinformatics and Computational Biology, University of Maryland, College Park MD 20742, USA.

5. “Using Bioinformatics Applications on the Cloud”, Hyungro Lee School of Informatics and Computing, Indiana University 815 E 10th St. Bloomington, IN 47408 lee212@indiana.edu.


7. https://Hadoop.apache.org/

**Index Terms**

| Computer Science | Security |

**Keywords**

Interest Locality, Data Groupings, CloudBLAST, Cloudburst