

Analysis of Genomics Big Data in Cloud Computing Using Fuzzy Logic

Authors : Mohammad Vahed, Ana Sadeghitohidi, Majid Vahed, Hiroki Takahashi

Abstract : In the genomics field, the huge amounts of data have produced by the next-generation sequencers (NGS). Data volumes are very rapidly growing, as it is postulated that more than one billion bases will be produced per year in 2020. The growth rate of produced data is much faster than Moore's law in computer technology. This makes it more difficult to deal with genomics data, such as storing data, searching information, and finding the hidden information. It is required to develop the analysis platform for genomics big data. Cloud computing newly developed enables us to deal with big data more efficiently. Hadoop is one of the frameworks distributed computing and relies upon the core of a Big Data as a Service (BDaaS). Although many services have adopted this technology, e.g. amazon, there are a few applications in the biology field. Here, we propose a new algorithm to more efficiently deal with the genomics big data, e.g. sequencing data. Our algorithm consists of two parts: First is that BDaaS is applied for handling the data more efficiently. Second is that the hybrid method of MapReduce and Fuzzy logic is applied for data processing. This step can be parallelized in implementation. Our algorithm has great potential in computational analysis of genomics big data, e.g. de novo genome assembly and sequence similarity search. We will discuss our algorithm and its feasibility.

Keywords : big data, fuzzy logic, MapReduce, Hadoop, cloud computing

Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development

Conference Location : Chicago, United States

Conference Dates : December 12-13, 2020