Bioinformatics High Performance Computation and Big Data

Authors : Javed Mohammed

Abstract : Right now, bio-medical infrastructure lags well behind the curve. Our healthcare system is dispersed and disjointed; medical records are a bit of a mess; and we do not yet have the capacity to store and process the crazy amounts of data coming our way from widespread whole-genome sequencing. And then there are privacy issues. Despite these infrastructure challenges, some researchers are plunging into bio medical Big Data now, in hopes of extracting new and actionable knowledge. They are doing delving into molecular-level data to discover bio markers that help classify patients based on their response to existing treatments; and pushing their results out to physicians in novel and creative ways. Computer scientists and bio medical researchers are able to transform data into models and simulations that will enable scientists for the first time to gain a profound under-standing of the deepest biological functions. Solving biological problems may require High-Performance Computing HPC due either to the massive parallel computation required to solve a particular problem or to algorithmic complexity that may range from difficult to intractable. Many problems involve seemingly well-behaved polynomial time algorithms (such as all-to-all comparisons) but have massive computational requirements due to the large data sets that must be analyzed. High-throughput techniques for DNA sequencing and analysis of gene expression have led to exponential growth in the amount of publicly available genomic data. With the increased availability of genomic data traditional database approaches are no longer sufficient for rapidly performing life science queries involving the fusion of data types. Computing systems are now so powerful it is possible for researchers to consider modeling the folding of a protein or even the simulation of an entire human body. This research paper emphasizes the computational biology's growing need for high-performance computing and Big Data. It illustrates this article's indispensability in meeting the scientific and engineering challenges of the twenty-first century, and how Protein Folding (the structure and function of proteins) and Phylogeny Reconstruction (evolutionary history of a group of genes) can use HPC that provides sufficient capability for evaluating or solving more limited but meaningful instances. This article also indicates solutions to optimization problems, and benefits Big Data and Computational Biology. The article illustrates the Current State-of-the-Art and Future-Generation Biology of HPC Computing with Big Data.

Keywords : high performance, big data, parallel computation, molecular data, computational biology

Conference Title : ICBCBBE 2014 : International Conference on Bioinformatics, Computational Biology and Biomedical Engineering

Conference Location : London, United Kingdom

Conference Dates : November 28-29, 2014