Genodata: The Human Genome Variation Using BigData

Authors : Surabhi Maiti, Prajakta Tamhankar, Prachi Uttam Mehta

Abstract : Since the accomplishment of the Human Genome Project, there has been an unparalled escalation in the sequencing of genomic data. This project has been the first major vault in the field of medical research, especially in genomics. This project won accolades by using a concept called Bigdata which was earlier, extensively used to gain value for business. Bigdata makes use of data sets which are generally in the form of files of size terabytes, petabytes, or exabytes and these data sets were traditionally used and managed using excel sheets and RDBMS. The voluminous data made the process tedious and time consuming and hence a stronger framework called Hadoop was introduced in the field of genetic sciences to make data processing faster and efficient. This paper focuses on using SPARK which is gaining momentum with the advancement of BigData technologies. Cloud Storage is an effective medium for storage of large data sets which is generated from the genetic research and the resultant sets produced from SPARK analysis.

Keywords : human genome project, Bigdata, genomic data, SPARK, cloud storage, Hadoop

Conference Title : ICDMD 2016 : International Conference on Data Mining and Database

Conference Location : Singapore, Singapore

Conference Dates : March 03-04, 2016