Healthcare Big Data Analytics Using Hadoop

Authors : Chellammal Surianarayanan

Abstract : Healthcare industry is generating large amounts of data driven by various needs such as record keeping, physician's prescription, medical imaging, sensor data, Electronic Patient Record(EPR), laboratory, pharmacy, etc. Healthcare data is so big and complex that they cannot be managed by conventional hardware and software. The complexity of healthcare big data arises from large volume of data, the velocity with which the data is accumulated and different varieties such as structured, semi-structured and unstructured nature of data. Despite the complexity of big data, if the trends and patterns that exist within the big data are uncovered and analyzed, higher quality healthcare at lower cost can be provided. Hadoop is an open source software framework for distributed processing of large data sets across clusters of commodity hardware using a simple programming model. The core components of Hadoop include Hadoop Distributed File System which offers way to store large amount of data across multiple machines and MapReduce which offers way to process large data sets with a parallel, distributed algorithm on a cluster. Hadoop ecosystem also includes various other tools such as Hive (a SQL-like query language), Pig (a higher level query language for MapReduce), Hbase(a columnar data store), etc. In this paper an analysis has been done as how healthcare big data can be processed and analyzed using Hadoop ecosystem.

Keywords : big data analytics, Hadoop, healthcare data, towards quality healthcare
Conference Title : ICACSIT 2015 : International Conference on Advanced Computer Science and Information Technology
Conference Location : Singapore, Singapore
Conference Dates : September 10-11, 2015