Hadoop is a [distributed storage](https://en.wikipedia.org/wiki/Clustered_file_system) and [distributed processing](https://en.wikipedia.org/wiki/Distributed_processing) of very large data sets on [computer clusters](https://en.wikipedia.org/wiki/Computer_cluster) . All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common and should be automatically handled by the framework.

The core of Apache Hadoop consists of a storage part, known as [Hadoop Distributed File System](https://en.wikipedia.org/wiki/Apache_Hadoop" \l "HDFS) (HDFS), and a processing part called [MapReduce](https://en.wikipedia.org/wiki/MapReduce" \o "MapReduce). Hadoop splits files into large blocks and distributes them across nodes in a cluster. To process data, Hadoop transfers [packaged code](https://en.wikipedia.org/wiki/JAR_(file_format)) for nodes to process in parallel based on the data that needs to be processed