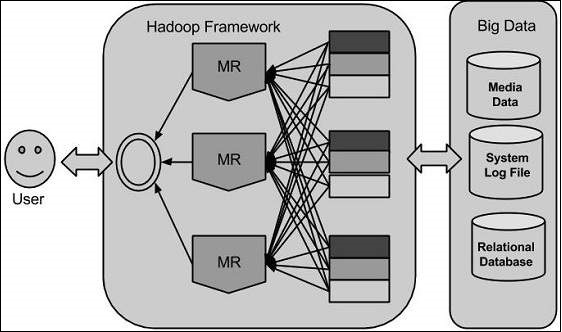
## Definition

* Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.
* Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel with others.
* Hadoop is used to develop applications that could perform complete statistical analysis on huge amounts of data.



Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

Hadoop has two major layers namely −

* Processing/Computation layer (MapReduce), and
* Storage layer (Hadoop Distributed File System).

Hadoop framework also includes the following two modules −

* Hadoop Common − These are Java libraries and utilities required by other Hadoop modules.
* Hadoop YARN − This is a framework for job scheduling and cluster resource management.

## Working

It is quite expensive to build bigger servers with heavy configurations that handle large scale processing, but as an alternative, one can tie together many commodity computers with single-CPU, as a single functional distributed system and practically, the clustered machines can read the dataset in parallel and provide a much higher throughput. Moreover, it is cheaper than one high-end server.

## Advantages of Hadoop

* Hadoop framework allows the user to quickly write and test distributed systems.
* Hadoop does not rely on hardware to provide fault-tolerance and high availability (FTHA), rather Hadoop library itself has been designed to detect and handle failures at the application layer.
* Servers can be added or removed from the cluster dynamically and Hadoop continues to operate without interruption.
* Another big advantage of Hadoop is that apart from being open source, it is compatible on all the platforms since it is Java based.