**COMPONENTS OF HADOOP 2.X:**

->YARN

->HDFS

->MAPREDUCE

I)YARN: (Yet Another Resource Negotiator )

* It is the key feature in hadoop 2.x, it is an extended version of mapreduce.
* It is a large-scale, distributed operating system for big data applications.
* It allows multiple data processing engines such as interactive SQL, real-time streaming, data science and batch processing to handle data stored in a single platform.
* The primary idea of YARN is to split up the functionalities of resource management and job scheduling/monitoring into separate daemons.
* That is to have a global resource manager (RM) and per application manager (AM).
* It separates the resource management and processing components.
* It is often used to separate the hdfs from mapreduce in order to make the hadoop environment more suitable for operational applications.
* Instead of jobtracker and tasktracker, there is a resource manager, application master and node manager.
* In yarn we can work with lot of cluster without any bottleneck issues.

II)HDFS: (Hadoop Distributed File System)

* It is a Big Data analytics tool which is used for storage .
* The file store in HDFS provides scalable, fault-tolerant storage at low cost.
* It stores files across the collection of servers in a cluster.
* It is used for storage purpose of data before processing and result after processing.
* It enables data transfer among the nodes.
* The HDFS software detects and compensates for hardware issues, including disk problems and server failure.
* Files are decomposed into blocks and each block is written to more than one of the servers.
* HDFS application is based on a write-once and read-many access model for files.
* A file once created, written, and closed need not be changed.

III)MAPREDUCE:

* MapReduce is a key processing technique concept in hadoop.
* The MapReduce algorithm contains two tasks, namely Map and Reduce.
* Map task takes a set of data and converts it into another set of data, where individual elements are broken down into key/value pair format.
* The reduce task on the other hand takes the output from a mapper class as its input and combines those into a smaller set of data.
* The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes.
* It delivers out work to various nodes within the map.
* It organizes and reduces the results from each node into a solution.