1.List the components of hadoop2.X and explain each component in detail.

* HDFS
* YARN
* Map Reduce
* HDFS:
* It is one of the core component of Hadoop used for the storage of large amounts of files.
* It provides scalablity, fault-tolerant storage of files.
* Reduces the cost required to store such huge amounts of data.
* The HDFS software detects and compensates for hardware issues, including disk problems and server failure.
* The files stored are decomposed into blocks which are then replicated on more than one of the servers.
* The replication provides both fault-tolerance and improved performance.
* It mainly consists of ->namenodes – stores metadata.

->datanodes - stores the data.

* Files are decomposed into blocks and each block is written to more than one of the servers.
* YARN:
* It is an abbreviated form of Yet Another Resource Negotiator.
* It is an extended or the second version of mapreduce in hadoop.
* It is referred to as the architectural center of Hadoop, the resource management framework that shares workloads on a single dataset.
* It provides the resource management while HDFS on the other hand is used to provide a scalable, fault-tolerant, cost-efficient storage for big data.
* It is often viewed as a large-scale distributed OS for many big data applications.
* It is often used to separate the hdfs from mapreduce inorder to make the hadoop environment more suitable for operational applications.
* Map Reduce:
* 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 tuples in key/value pairs format.
* The reduce task on the other hand takes the output from a mapper class as its input and combines those data tuples into a smaller set of tuples.
* The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes.