**List the component of Hadoop 2.x and explain each component in detail.**

The Major 3 components of Hadoop2.x are .

1> HDFS

2>YARN

3>MapReduce

4> Hadoop Common Module

**In-Detail**

1> HDFS (Hadoop Distribution File System) Its is used as Distribution Storage System in Hadoop Architecture.

HDFS Cluster primarily consists of NameNode that manages file system metadata and

DataNode that stores the actual data.

* Hadoop, including HDFS, is well suited for distributed storage and distributed processing using commodity hardware. It is fault tolerant, scalable, and extremely simple to expand. MapReduce, well known for its simplicity and applicability for large set of distributed applications, is an integral part of Hadoop.
* HDFS is highly configurable with a default configuration well suited for many installations. Most of the time, configuration needs to be tuned only for very large clusters.
* Hadoop is written in Java and is supported on all major platforms.
* Hadoop supports shell-like commands to interact with HDFS directly.
* The NameNode and Datanodes have built in web servers that makes it easy to check current status of the cluster.

**Namenode**: Namenode is the heart of the hadoop system. The namenode manages the file system namespace. It stores the metadata information of the data blocks. This metadata is stored permanently on to local disk in the form of namespace image and edit log file.

**Secondary Namenode**: The responsibility of secondary name node is to periodically copy and merge the namespace image and edit log. In case if the name node crashes, then the namespace image stored in secondary namenode can be used to restart the namenode.

**DataNode**: It stores the blocks of data and retrieves them. The datanodes also reports the blocks information to the namenode periodically.

**JobTracker**: JobTracker responsibility is to schedule the clients jobs. Job tracker creates map and reduce tasks and schedules them to run on the datanodes (tasktrackers). Job Tracker also checks for any failed tasks and reschedules the failed tasks on another datanode. Jobtracker can be run on the namenode or a separate node.

**TaskTracker**: Tasktracker runs on the datanodes. Task trackers responsibility is to run the the map or reduce tasks assigned by the namenode and to report the status of the tasks to the namenode.