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

**Answer:**

Hadoop 2.x has below components:

1. **Hadoop Distributed File System(HDFS)**

The Hadoop Distributed File System ( **HDFS** ) is a distributed file system designed to run on commodity hardware

1. **MapReduce:**

A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system

1. **Yet Another Resource Negotiator(YARN):**

Apache Hadoop YARN (Yet Another Resource Negotiator) is a cluster management technology. YARN is one of the key features in the second-generation Hadoop 2 version of the Apache Software Foundation's open source distributed processing framework. It has below daemons:

* **ResourceManager**

This daemon process runs on master node. It is responsible for getting job submitted from client and schedule it on cluster, monitoring running jobs on cluster and allocating proper resources on the slave node. It communicates with Node Manager daemon process on the slave node to track the resource utilization. It uses two other processes named *Application Manager* and *Scheduler* for MapReduce task and resource management.

1. ApplicationManager -The Application Manager oversees the full lifecycle of an application, all the way from requesting the needed containers from the Resource Manager to submitting container lease requests to the NodeManager.
2. Scheduler – Responsible to schedule required resources to Applications (that is Per-Application Master). It does only scheduling. It does care about monitoring or tracking of those Applications.

* **NodeManager:**

Node Manager is a Per-Node Level component. It is responsible for managing the life-cycle of the Container and monitoring each Container’s Resources utilization. **Node Manager** (many per cluster) is the slave of the infrastructure. When it starts, it announces himself to the Resource Manager. Periodically, it sends a heartbeat to the Resource Manager. Each Node Manager offers some resources to the cluster. Its resource capacity is the amount of memory and the number of vcores. At run-time, the Resource Scheduler will decide how to use this capacity.

* **Container:**

Container is a fraction of the Namenode’s capacity and it is used by the client for running a program.