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

Hadoop 2x added additional features to HDFS and introduced YARN

**YARN:**

YARN framework replaced the job responsibilities of job tracker and task tracker with separate components Global Resource Manager ,Node Manager , Application specific Application Master, Schedular, Container.

**Global** **Resource manager:**

* Global Resource manager is per cluster level.
* Resource manager is again divided into two components:
* - Application Manager

- Schedular

* **Resource managers scheduler** is responsible for scheduling resources to applicationsthat is per – application master.
* Resource managers scheduler does only scheduling, it does not care about monitoring or tracking those applications

**Application Master**

* Application Manager is per application level component
* It is responsible for managing assigned application life cycle
* It interacts with both resource manager scheduler and node manager.
* It interacts with resource managers scheduler to acquire required resources
* It interacts with node manager to execute assigned tasks and monitor those tasks
* It is the actual instance which does processing . It requests resource manager for resource allocation and works with node manager to get those resources for job execution. Application Master could be map reduce or any other framework.

**Node Manager**

* Node Manager is per node level component
* It is responsible for managing life cycle of the container
* It monitors each containers resource utilization
* It runs of each nodes and communicates to resource manager about resource utilization on machine
* It receives requests from resource manager about resource allocation to jobs and maintains the life cycle of containers

**Container**

* Each master node and slave node contains container
* Container is a portion of memory in HDFS(Either Name node or Data node)
* It is a set of allocated system resources like memory , CPU .

Resource is handled by Resource manager and Node manger

Processing is handled by Application Master.

**HDFS** has undergone a major transformation with a collection of new features that include:

* **NameNode HA**: Automated failover with a hot standby and resiliency for the NameNode master service.
* **Snapshots**: point-in-time recovery for backup, disaster recovery and protection against use errors.
* **Federation**: It is a clear separation of namespace and storage by enabling generic block storage layer.In simple words it allowd having more than one name node which shares the common storage pool