**Problem Statement**

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

1. **Components of Hadoop 2.x [YARN- Yet Another Resource Negotiator]**

* **Resource Manager –** assigns resource among applications for optimal resource utilization. One Cluster has one instance of Resource Manager.
* **Node Manager-** Runs on each node and communicates with resource manager about resource usage on the machine. It receives request from resource manager about resource allocation to jobs and maintains life cycle of containers.
* **Application Master-** It is the actual instance which does processing. It request Resource Manager for resources and works with Node Manager to get those resource for task execution. Application master could be Map Reduce or any other processing framework.
* **Scheduler-** It is plugged with resource Manager to help in resource allocation. Different schedulers allocate resources using different algorithm.
* **Container-** It is a set of allocated system resources (CPU Core and Memory). Containers are allocated and managed by Node Manager and are used by tasks.

Resource is handled by resource manager and node manager.

Processing is handled by application Master (Map Reduce is one of the many possible types of application Master). Processing other than Map Reduce is also possible.





* **Step1:** Job/Application (which can be Map Reduce, Java/Scala Application, DAG jobs like Apache Spark etc)is submitted by the YARN client application to the Resource Manager daemon along with the command to start the Application Master on any container at Node Manager.
* **Step2:** Application Manager process on Master Node validates the job submission request and hands It over to scheduler process for resource allocation.
* **Step3:** Scheduler process assigns a container for Application Master on one slave node.
* **Step4:** Node Manager daemon starts the Application Master service within one of its container using the command mentioned in step 1, hence Application Master is considered to be the first container of any application.
* **Step5:** Application Master negotiates the other containers from Resource Manager by providing the details like location of data on slave nodes, required CPU, memory , cores etc.
* **Step 6:** Resource Manager allocates the best suitable resources on slavenodes and responds to Application master with node details and other details.
* **Step 7:** Then, Application Master send request to Node Managers on suggested slave nodes to start the containers.
* **Step 8:** Application master then manages the resources of requested containers while job execution and notifies the Resource Manager when execution is completed.
* **Step 9:** Node Manager periodically notify the Resource Manager with the current status of available resources on the node as to what information can be used by scheduler to schedule new application on the clusters.
* **Step 10:** In case of any failure of slave node, Resource manager will try to allocate new container on other best suitable node so that application Master can complete the process using new container.