**ASSIGNMENT 3.1**

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

Solutions:

Hadoop 2.x consist of the following components:

(i) Yarn:

YARN is a software rewrite that decouples MapReduce's resource management and scheduling capabilities from the data processing component, enabling Hadoop to support more varied processing approaches and a broader array of applications.

YARN combines a central resource manager that reconciles the way applications use Hadoop system resources with node manager agents that monitor the processing operations of individual cluster nodes. Running on commodity hardware clusters, Hadoop has attracted particular interest as a staging area and data store for large volumes of structured and unstructured data intended for use in analytics applications.

(ii) Resource Manager:

* Resource Manager is a Per-Cluster Level Component.
* Resource Manager is again divided into two components:
  1. Scheduler
  2. Application Manager
* Resource Manager’s Scheduler is :
  1. Responsible to schedule required resources to Applications (that is Per-Application Master).
  2. It does only scheduling. It does care about monitoring or tracking of those Applications.

(iii) Application Master:

Application Master is a per-application level component. It is responsible for:

* 1. Managing assigned Application Life cycle.
  2. It interacts with both Resource Manager’s Scheduler and Node Manager
  3. It interacts with Scheduler to acquire required resources.
  4. It interacts with Node Manager to execute assigned tasks and monitor those task’s status.

(iv) Node Manager:

* Node Manager is a Per-Node Level component.
* It is responsible for:
  1. Managing the life-cycle of the Container.
  2. Monitoring each Container’s Resources utilization.

(v) Schedulers:

It is plugged with Resource Manager to help in resource allocation. Different schedulers allocate resources using different algorithms.

(vi) Containers:

It is a set of allocated system resources (CPU Core and Memory). Containers are allocated and managed by NodeManager and are used by tasks.