**ASSIGNMENT-3.3**

1. Components of Hadoop 2.x:

Hadoop 2.x has the following three Major Components:

* HDFS
* YARN
* Map Reduce

***HDFS:***

* HDFS stands for Hadoop Distributed File System.
* It is also known as HDFS V2 as it is part of Hadoop 2.x with some enhanced features.
* It is used as a Distributed Storage System in Hadoop Architecture.

***YARN:***

* YARN stands for Yet Another Resource Negotiator.
* It is new Component in Hadoop 2.x. It is also known as “MR V2”.
* It is a software rewrite that decouples MapReduce's resource management and scheduling capabilities from the data processing component, enabling [Hadoop](http://searchcloudcomputing.techtarget.com/definition/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](http://whatis.techtarget.com/definition/software-agent) that monitor the processing operations of individual cluster nodes

***MapReduce***:

* Map Reduce is a Batch Processing or Distributed Data Processing Module.
* It is also known as “MR V1” as it is part of Hadoop 1.x with some updated features.

***The updated features are to overcome some problems in HADOOP 1.x***

***Problems in Hadoop 1 .x:***

1. Name Node is a single point of failure

***Solution:*** High Availability: In HDFS High Availability, multiple Name Nodes are used in the Active-Standby mode with shared edits to handle the Name Node failure.

1. Secondary Name Node cannot act as Name Node

***Solution:***

HDFS Federation: Hadoop 2.x Architecture which allows to manage multiple namespaces by enabling multiple Name Nodes. So on HDFS shell you have multiple directories available but it may be possible that two different directories are managed by two active Name Nodes at a time.