Hadoop 2.x:

Components:

1. Name Node
2. Secondary Name Node
3. Data Node
4. YARN (Resource Manager , Node Manager, Application-specific application master, Containers & Schedulers)
5. Name Node:
   * Contains Hadoop FileSystem Tree and other metadata information about files and directories.
   * In-Memory mapping of which blocks are stored in which datanode.
6. Secondary Namenode:

• Performs house-keeping activities for namenode, like periodic merging of namespace and edits.

• This is not a backup for namenode.

1. DataNode:

• Stores actual data blocks of file in HDFS on its own local disk.

• Sends signals to NameNode periodically (called as Heartbeat) to verify it is active.

• Sends block reporting to the nameode on cluster startup as well as periodically at every 10th Heartbeat.

• The data node are the workhorse of the system.

• They perform all the block operation including periodic checksum. They receive instructions from the name node of where to put the blocks and how to put the blocks.

1. YARN: Yet Another Resource Negotiator
   1. Resource Manager:
      1. Assigns resources among applications for optimal resource utilization.
      2. One cluster has one instance of Resource Manager.
   2. Node Manager:
      1. Runs on each node and communicates with Resource Manager about resource usage on the machine.
      2. It receives requests from resource manager about resource allocation to jobs and maintains life cycle of containers.
   3. Application-specific Application Master:
2. It is the actual instance which does processing.
3. It requests Resource Manager for resources and works with NodeManager to get those resources for task execution. Application Master could be MapReduce or anyother processing framework.
   1. Scheduler:
4. It is plugged with Resource Manager to help in resource allocation.
5. Different Schedulers allocate resources using different algorithms.
   1. Container:
      1. It is a set of allocated system resources (CPU Core and Memory).
      2. Containers are allocated and managed by NodeManager and are used by tasks.