Daemons is a process that runs in the background. NameNode, DataNode, Resource manager, Node manager are part of hadoop daemons running on hadoop Generation-II. Each daemons runs separately in its own JVM. Importance of each daemon is explained below

Name Node:

* There is only single instance of this process runs on a cluster and that is on a master node
* It is responsible for manage metadata about files distributed across the cluster
* It manages information like location of file blocks across cluster and it’s permission
* This process reads all the metadata from a file named fsimage and keeps it in memory
* After this process is started, it updates metadata for newly added or removed files in RAM
* It periodically writes the changes in one file called edits as edit logs
* This process is a heart of HDFS, if it is down HDFS is not accessible any more

DataNode :

* There are many instances of this process running on various slave nodes(referred as Data nodes)
* It is responsible for storing the individual file blocks on the slave nodes in Hadoop cluster
* Based on the replication factor, a single block is replicated in multiple slave nodes(only if replication factor is > 1) to prevent the data loss
* Whenever required, this process handles the access to a data block by communicating with Name Node
* This process periodically sends heart bits to Name Node to make Name Node aware that slave process is running
* The DataNode sends a heartbeat message to the NameNode every 3 seconds to inform the NameNode that it is alive. If the NameNode doesn’t receive a heartbeat message from the DataNode in 10 mins (configurable), it considers the DataNode to be dead. It then stores the replica of the block in some other DataNode.
* The Client receives an ACK form the DataNode that it has received the data. If it doesn’t after several tries, it is understood that either there is network failure or the DataNode has failed.
* Checksum is sent along with the data to look for data corruption.
* Periodically the DataNodes sends the report containing the list of blocks that are uncorrupted. The NameNode then updates the list of valid blocks a DataNode contains.
* For all such under replicated blocks, the NameNode adds other DataNodes to the replication pipeline.

Resource Manager

* This daemon process runs on master node (may run on the same machine as name node for smaller clusters)
* It is responsible for getting job submitted from client and schedule it on cluster, monitoring running jobs on cluster and allocating proper resources on the slave node
* It communicates with Node Manager daemon process on the slave node to track the resource utilization
* It uses two other processes named Application Manager and Scheduler for MapReduce task and resource management

Node Manager

* This daemon process runs on slave nodes (normally on HDFS Data node machines)
* It is responsible for coordinating with Resource Manager for task scheduling and tracking the resource utilization on the slave node
* It also reports the resource utilization back to the Resource Manager
* It uses other daemon process like Application Master and Container for MapReduce task scheduling and execution on the slave node