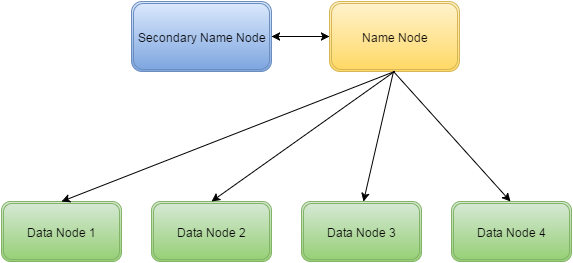
**Components of Hadoop 1.x:**

Hadoop System is a Master-Slave architecture. Hadoop 1.x file system has 64 MB block size (We can change it as per our Project requirements).The default replication factor in Hadoop is 3, which is configurable.

There are total 5 components in Hadoop 1.x architecture:

1. Name Node   
2. Data Node   
3. Secondary Name Node   
4. Job Tracker   
5. Task Tracker

* Name Node, Data Node and Secondary Name Node - Storage components.
* Job Tracker and Task Tracker - Processing components.



**NameNode:**

* Name node is a Master Node in Hadoop.
* The job of name node is to decide, how to store the physical location of each and every file in blocks in the cluster
* Contains the Hadoop File System Tree and other metadata information about files and directories.
* Contains in-memory mapping of which blocks are stored in which datanode.
* Namenode always stores the Metadata in FSImage and EditLogs file at regular intervals. This process is called as Checkpoint mechanism

***Note:*** The Name node is meant for maintaining the metadata information of the complete hadoop cluster, but it will never stores the actual data. The Actual data will always be stored in the Data node only.

**Secondary NameNode:**

* Main function is to take checkpoints of the file system metadata present on namenode.
* Performs house-keeping activities for NameNodes, like the periodic merging of namespace and edits.
* This is not a backup for a NameNode.

**DataNode:**

* Stores actual data blocks of files in HDFS on its own local disk.
* Data node will send the RPC signal at regular interval, to notify the Name node that it is alive and working.(called as Heart beat mechanism)
* Sends block reporting to the NameNode on the cluster startup as well as periodically at every 10th Heartbeat.
* The DataNodes are the workhorses of a system.
* They perform all the block operations including periodic checksum. They receive instructions from the name node of where to put the blocks and how to put them.

**Job Tracker:**

* Controls the overall execution of the MapReduce jobs.
* Its job is to assign the task to the Data nodes/Task trackers.
* It also decides the job scheduling for the data nodes/Task trackers.
* In case of Job failure, Job tracker decides about the rescheduling of the task on some other nodes.

**Task Tracker:**

* Task tracker's job is to execute the task assigned by Job tracker.
* Runs individual MapReduce jobs on DataNodes  
  periodically communicates with the Job Tracker to give updates and receive instructions.
* The [Task Tracker](https://wiki.apache.org/hadoop/TaskTracker) nodes are monitored. If they do not submit heartbeat signals often enough, they are deemed to have failed and the work is scheduled on a different [Task Tracker](https://wiki.apache.org/hadoop/TaskTracker).

