**HDFS** (storage) and **MapReduce** (processing) are the two core components of Apache Hadoop.

The main components of HDFS are as described below:

• **NameNode** is the master of the system. It maintains the name system (directories and files) and manages the blocks which are present on the DataNodes.

• **DataNodes** are deployed on each machine and provide the actual storage. They are responsible for serving read and write requests for the clients.

• Secondary NameNode is responsible for performing periodic checkpoints. In the event of NameNode failure, the NameNode can be restarted using the checkpoint.

The main components of **MapReduce** are as described below:

• **JobTracker** is the master of the system which manages the jobs and resources in the cluster (TaskTrackers). The JobTracker tries to schedule each map as close to the actual data being processed i.e. on the TaskTracker which is running on the same DataNode as the underlying block.

• **TaskTrackers** are deployed on each machine. They are responsible for running the map and reduce tasks as instructed by the JobTracker.

• **JobHistoryServer** is a daemon that serves historical information about completed applications. Typically, JobHistory server can be co-deployed with JobTracker, but it is recommended to run as a separate daemon.