With growing data velocity the data size easily outgrows the storage limit of a machine. A solution would be to store the data across a network of machines. Such filesystems are called *distributed filesystems*. Since data is stored across a network all the complications of a network come in.   
This is where Hadoop comes in. It provides one of the most reliable filesystems. HDFS (Hadoop Distributed File System) is a unique design that provides storage for *extremely large files* with streaming data access pattern and it runs on *commodity hardware*. Let’s elaborate the terms:

* ***Extremely large files***: Here we are talking about the data in range of petabytes(1000 TB).
* ***Streaming Data Access Pattern***: HDFS is designed on principle of *write-once and read-many-times*. Once data is written large portions of dataset can be processed any number times.
* ***Commodity hardware:***Hardware that is inexpensive and easily available in the market. This is one of feature which specially distinguishes HDFS from other file system.

**Nodes:** Master-slave nodes typically forms the HDFS cluster.

1. **NameNode(MasterNode):**
   * Manages all the slave nodes and assign work to them.
   * It executes filesystem namespace operations like opening, closing, renaming files and directories.
   * It should be deployed on reliable hardware which has the high config. not on commodity hardware.
2. **DataNode(SlaveNode):**
   * Actual worker nodes, who do the actual work like reading, writing, processing etc.
   * They also perform creation, deletion, and replication upon instruction from the master.
   * They can be deployed on commodity hardware.

**HDFS daemons:** Daemons are the processes running in background.

* **Namenodes:**
  + Run on the master node.
  + Store metadata (data about data) like file path, the number of blocks, block Ids. etc.
  + Require high amount of RAM.
  + Store meta-data in RAM for fast retrieval i.e to reduce seek time. Though a persistent copy of it is kept on disk.
* **DataNodes:**
  + Run on slave nodes.
  + Require high memory as data is actually stored here.

