**Hadoop**

Hadoop is an application/library/service? which runs on the Linux operating system.

It provides a bunch of services, the most important ones being:

1. HDFS or Hadoop distributed file system.

2. Map/Reduce

The most important concept to understand here is that, HDFS and Map/Reduce are just some services which Hadoop keeps running when we kick off the Hadoop application.

So the Linux OS on which we're running Hadoop, doesn't know anything about the distributed nature of the filesystem (HDFS) and all the other computers that Hadoop it's connected to form the Hadoop cluster. This is all managed through Hadoop.

So, HDFS is an application filesystem. And application filesystems can be seen only by their respective applications and are not seen by and cannot be accessed by the host OS on which the application runs.

HDFS is part of the Hadoop library or application and it can't be accessed through the local filesystem of the host computer, ext4 or ext3 is the filesystem type on Linux and HDFS will not show up as folder or directory structure on this Linux filesystem.

Apart from a general understanding of Hadoop this piece of information is pertinent to filesystem functions.

Like say we want to perform some functions like get a file's date or set a file's date etc while we use HDFS for stroing our files, then we can't use java.io (java) fs (clojure) modules for performing these functions.

That's because these libraries are meant to work with the OS filesystem. So, when I use java.io or fs in clojure to obtain the file date, then the function calls would be making calls to the OS, where then the OS would be looking for files on it's filesytem based on the path specified. And since the OS can't see HDFS, a file on HDFS won't be found by these functions.

*— MAR 30, 2015 —*