

## Hadoop

En palabras fáciles Hadoop puede verse como un set de programas open source (son libres para que cualquier persona los pueda utilizar o modificar) que cualquier persona puede usar como la “columna vertebral” de su aplicación de big data.

### Los 4 módulos de Hadoop

Hadoop esta formado por módulos los cuales hacen tareas esenciales para un sistema diseñado para Big Data Analytics.

#### **Distributed File-System:**

Los módulos más importantes son DFS, que permite almacenar datos en un formato muy accesible a través de una gran cantidad de dispositivos de almacenamiento unidos y MapReduce que provee las herramientas básicas para poder hacer experimentos y operaciones con los datos.

Un “file-system” es el método usado por las computadoras para almacenar datos, y que puedan ser encontrados y accedidos. Normalmente se determina por el Sistema Operativo, sin embargo Hadoop usa su propio sistema FS que se posa sobre el FS de la computadora.

#### **MapReduce:**

Es llamado por las dos operaciones básicas que este módulo realiza leer datos de la base datos, poner en un formato listo para análisis (map) y realizar operaciones matemáticas, por ejemplo: contar el número de hombres mayores a 30 años en una base de datos de clientes (reduce).

#### **Hadoop Common:**

Este modulo provee las herramientas (en Java) necesitadas por el sistema del usuario para leer los datos guardados bajo el file system de Hadoop.

#### **YARN:**

El último modulo es YARN, el cual administra los recursos de los sistemas que almacenan los datos y corren los análisis.

Hay más programas que han llegado a formar parte de Hadoop, pero los 4 antes mencionados son los considerados core de la plataforma.

### Cómo nace Hadoop

El desarrollo de Hadoop inició cuando unos ingenieros de software se dieron cuenta que muchos sistemas almacenando y analizando en paralelo eran más rápidos que uno grande.

Esto en parte porque mientras más grande sea el dispositivo de almacenamiento, más tiempo toma para leer, encontrar información y moverla a un segmento determinado.

El código fue liberado en 2005 por la “Apache Software Foundation”, una organización sin fines de lucro que produce software libre. El nombre de Hadoop era el nombre de un elefante de peluche que tenía el hijo de uno de los creadores originales.

**Sitio:** [hadoop.apache.org](http://hadoop.apache.org)

**Tutorial:** <https://www.tutorialspoint.com/hadoop/index.htm>