

Pérez Ruiz Miguel Ángel

Hadoop Distributed File System

Es distribuido ya que nos muestra una representación más significativa de este sistema de ficheros y además que es su capacidad para almacenar los archivos en un clúster de varias máquinas.

Nos permite el procesamiento distribuido de grandes cantidades de datos usando modelos de programación simple sobre un clúster.

- HDFS, es un sistema de ficheros que está especialmente diseñado para funcionar bien cuando se almacenan archivos grandes, que posteriormente se leerán de forma secuencial.
- HDFS no se comporta especialmente bien cuando lo que se pretende es realizar accesos aleatorios a los archivos, ni cuando estos se actualizan frecuentemente.
- Como se ha comentado, el hecho de que el sistema de ficheros sea distribuido proporciona ventajas, pues se pueden almacenar muchos más datos que los que se podrían almacenar en una sola máquina. Además, en general la cantidad de datos que se puede almacenar puede escalar con facilidad, ya que basta con añadir nuevos nodos al clúster para aumentar su capacidad de almacenamiento.
- HDFS proporciona redundancia, es decir, almacena los ficheros varias veces y en varios equipos distintos, para evitar que, si uno de ellos falla, los datos se pierdan.

Las características de HADOOP son:

Procesamiento distribuido

Eficiente

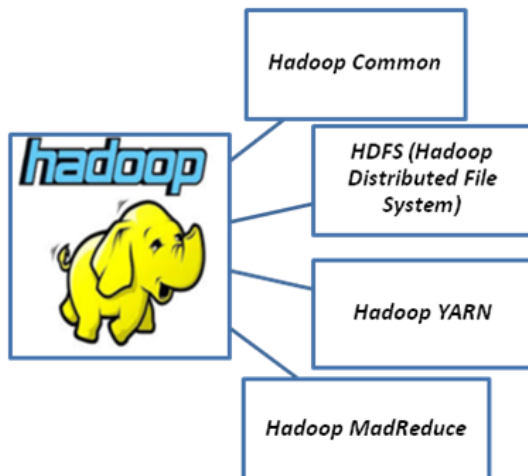
Económico

Fácilmente escalable

Tolerante a fallos

Open Source

Arquitectura básica:



Hadoop Common: Lo conforman las librerías para formar que son necesarias para ejecutar hadoop.

HDFS: Sistema de archivos de hadoop distribuido

YARN: Gestor de recursos de hadoop.

MapReduce: Manera de implementar la paralelización de los datos

Algunas opciones para trabajar con hadoop son:

Microsoft Azure, Horton Works, Cloudera.

CIBERGRAFÍA:

https://www.ibm.com/support/knowledgecenter/es/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginsights.product.doc/doc/c0057606.html

https://es.wikipedia.org/wiki/Hadoop_Distributed_File_System

<https://www.youtube.com/watch?v=yJsITGEmWm8>