

2

HDFS → basado en GPS : Google File System  
MapReduce

!lo va a preguntar

# T2 : Hadoop

## Conceptos Generales

Hadoop → Framework de código abierto

✓ Utilizado para procesamiento distribuido de datos grandes en clusters  
✓ Almacenamiento y procesamiento de datos grandes en clusters

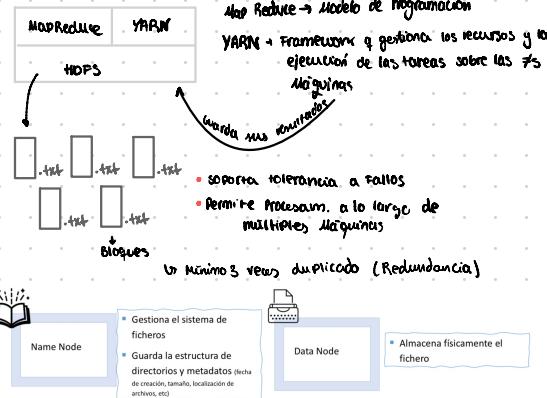
sistema de Archivos distribuidos → Hadoop Distributed File System (HDFS)

Modelo de Programación MapReduce

HDFS → + los archivos en bloques

distribuye  
nodos  
nodos  
nodos de un cluster  
Almacenamiento escalable redundante

Componentes Hadoop



## Comandos Básicos

```
hdfs dfs -copyFromLocal <fileorigen> <directorioHadoop>  
hdfs dfs -put <fileorigen> <directorioHadoop>  
  
hdfs dfs -appendToFile <fileorigen> <fileHadoop>  
  
hdfs dfs -moveFromLocal <fileorigen> <directorioHadoop>  
  
hdfs dfs -get <fileHadoop> <directorioLocal>  
hdfs dfs -copyToLocal <fileHadoop> <filelocal>  
  
hdfs dfs -cp origen destino  
hdfs dfs -mkdir <directorio>  
hdfs dfs -mv origen destino  
hdfs dfs -rm -r <file> <directorio>  
hdfs dfs -ls <directorio>  
hdfs dfs -ls <directorio> -- patron  
hdfs dfs -ls -R <directorio> --lectura en formato amigable  
hdfs dfs -ls -R <directorio> --lectura recursiva  
  
hdfs dfs -chmod <-R> XXX fichero  
hdfs dfs -chown <-R> usuario:grupo fichero <directorio> -- cambio del owner  
hdfs dfs -chgrp <-R> grupo fichero <directorio> -- cambio de grupo  
  
hdfs dfs text fichero → salida en formato legible  
hdfs dfs -d cat fichero → salida del fichero as is
```

!hdfs dfs -put: copiar en HDFS  
!hdfs dfs -ls: mostrar fichero  
!hdfs dfs -lsj: mostrar ult líneas del archivo  
hadoop fs -ls: (de csv) muestra primera línea del archivo  
  
!ls -t -n → muestra tamaño  
!ls -t por Fecha y hora de mod.

## Creación de un directorio:

```
hdfs dfs -mkdir user/datos/pruebas/claseprueba  
# Si lo ejecuta en el terminal... → se crea el directorio  
# Si se ha creado en el sistema hdfs  
  
hdfs dfs -mkdir user/datos/pruebas/claseprueba  
user/datos/pruebas/claseprueba  
  
hdfs dfs -ls  
Found 2 items  
drwxr-xr-x - root hadoop 0 2024-01-08 12:37 user/datos/pruebas/claseprueba  
drwxr-xr-x - root hadoop 0 2024-01-08 12:39 claseprueba  
  
hdfs dfs -ls  
Found 2 items  
drwxr-xr-x - root hadoop 0 2024-01-08 12:37 user/datos/pruebas/claseprueba  
drwxr-xr-x - root hadoop 0 2024-01-08 12:39 claseprueba  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 user/datos/pruebas/archivo.txt  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 user/datos/pruebas/archivo2_2099.csv  
  
hdfs dfs -ls /user/datos/pruebas/claseprueba  
Found 2 items  
drwxr-xr-x - root hdfs 0 2024-01-08 12:37 user/datos/pruebas/claseprueba  
drwxr-xr-x - root hdfs 0 2024-01-08 12:39 claseprueba
```

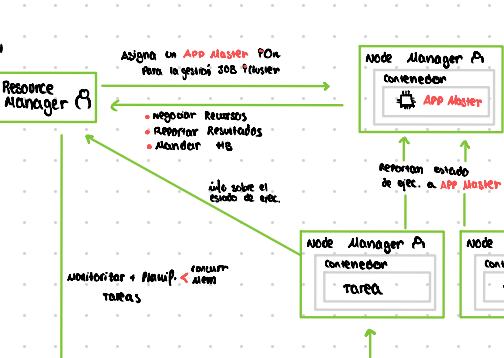
## Copiado de Archivos:

```
hdfs dfs -put archivo.txt /user/datos/pruebas # Se COPIA 'archivo.txt' en hdfs 'user'  
  
hdfs dfs -ls /user/datos/pruebas  
Found 1 items  
drwxr-xr-x 2 root hadoop 0 2024-01-08 12:34 /user/datos/pruebas/archivo.txt  
  
hdfs dfs -copyFromLocal /home/datos/pruebas/archivo2_2099.csv /user/datos/pruebas  
  
hdfs dfs -ls /user/datos/pruebas  
Found 2 items  
drwxr-xr-x - root hdfs 0 2024-01-08 12:37 user/datos/pruebas/claseprueba  
drwxr-xr-x - root hdfs 0 2024-01-08 12:39 claseprueba  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 user/datos/pruebas/archivo.txt  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 user/datos/pruebas/archivo2_2099.csv
```

## listado de Archivos en Hadoop:

```
hdfs dfs -ls /  
Found 4 items  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/hadoop  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/hdfs  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/kafka  
  
hdfs dfs -ls /user  
Found 13 items  
drwxr-xr-x - root hdfs 0 2024-01-08 12:31 /user/datos  
drwxr-xr-x - root hdfs 0 2024-01-08 12:31 /user/hadoop  
drwxr-xr-x - root hdfs 0 2024-01-08 12:31 /user/hdfs  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas2  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas3  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas4  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas5  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas6  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas7  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas8  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas9  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas10  
drwxr-xr-x - root hdfs 0 2024-01-08 12:35 /user/pruebas11
```

## YARN: Yet Another Resource Negotiator



## (Cantidad delegada) gestión en jobs

- Monitorea el estado de los ejec.
- Negocia Recursos con RMIA
- Reporta Resultados al RMIA
- Manda Job al RMIA

