

# Práctica 2 · WordCount sobre *El Quijote* (HDFS + MapReduce + YARN)

## Objetivo

- Cargar *quijote.txt* en HDFS.
- Ejecutar WordCount (MapReduce).
- Revisar salida y métricas en YARN/HistoryServer.

## 1) Preparación rápida

1. Descarga *quijote.txt* en tu VM (p. ej., en ~/Descargas).
2. Sube a HDFS:
  3. `hdfs dfs -mkdir -p /practicas`
  4. `hdfs dfs -put ~/Descargas/quijote.txt /practicas/`
  5. `hdfs dfs -ls /practicas`
6. (Si usas Hadoop 3.x y falla el classpath) añade `yarn.application.classpath` en `yarn-site.xml` y reinicia YARN.
7. (Si mata contenedores por memoria 143) añade `yarn.nodemanager.vmem-check-enabled=false` y reinicia YARN.

Servicios necesarios en marcha: HDFS, YARN y JobHistoryServer.

## 2) Ejecutar el job

- Elimina la salida si existe:
- `hdfs dfs -rm -r -f /practicas/resultado`
- Lanza WordCount con el jar de ejemplos (ajusta versión/ruta):
  - `hadoop jar .../hadoop-mapreduce-examples-<versión>.jar wordcount \ /practicas/quijote.txt /practicas/resultado`
  - Observa el progreso y el tracking URL.

## 3) Comprobar resultados

- Lista y visualiza:
- `hdfs dfs -ls /practicas/resultado`
- `hdfs dfs -cat /practicas/resultado/part-r-00000 | head`

- (Opcional) Baja el resultado:
- `hdfs dfs -get /practicas/resultado/part-r-00000 /tmp/palabras_quijote.txt`

#### 4) Analizar en la web

- YARN RM UI: `http://<rm>:8088/cluster` → tu aplicación → estado, tiempos.
- HistoryServer: `http://<hs>:19888/` → Job Overview → Maps/Reduces, Counters, Map tasks, Reduce tasks (shuffle/merge/reduce).
- Abre un map o reduce concreto y apunta: nodo, duración, reintentos.

#### 5) Entregables

1. Captura de /practicas en HDFS con quijote.txt.
2. Captura de /practicas/resultado mostrando \_SUCCESS y part-r-00000.
3. 10 primeras líneas del resultado.
4. Captura del Job Overview (YARN/History).
5. Captura de Map tasks y otra de Reduce tasks.
6. Respuestas breves:
  - Nº de map tasks y reduce tasks.
  - HDFS bytes read y HDFS bytes written.
  - ¿Qué fase tardó más (map/shuffle/reduce) y por qué crees que fue así?
  - ¿Hubo reintentos? ¿En qué tarea?