
BIG DAT-HADOOP

PROCESOS MAPREDUCE

Map Reduce

- La mayor parte de las herramientas de consulta están diseñadas para realizar consultas simples que deben ejecutarse rápidamente.
- El dato suele estar indexado y por tanto solo pequeñas porciones de datos se examinan durante la búsqueda.
- Esta solución, en cambio no es útil para datos no indexados de tipo semi estructurado (textos) o sin estructurar (multimedia).
- Para responder una query en esta solución es necesario examinar todos los datos
- Hadoop utiliza MapReduce para realizar un análisis exhaustivo de forma rápida.

Map Reduce

- **MapReduce**
- Es un algoritmo de procesamiento de datos que implementa un proceso en paralelo
- De forma simple distribuye las tareas a través de los nodos de un cluster ejecutando una función “map”.
 - La función map estudia el problema, lo divide en trozos y los manda a diferentes máquinas para que todos los trozos puedan ejecutarse concurrentemente..
 - Los resultados de este proceso paralelo se recogen y se distribuyen a través de un distintos servidores que ejecutan una función “reduce”, que toma los resultados de los trozos y los recombina para obtener una respuesta simple

Map Reduce

- **Procesos Mapper:**

- Los programas Map Reduce se dividen en Mappers, tareas que se ejecutan en los nodos
- Cada tarea MAP ataca a un solo bloque de datos HDFS
- Se ejecuta en el nodo donde reside el bloque (salvo excepciones)

- **Shuffle y Sort**

- Ordena y consolida los datos intermedios (temporales) que generan los mappers
- Se lanzan después de que todos los mappers hayan terminado y antes de que se lancen los procesos Reducer

- **Reducer**

- Operan sobre los datos intermedios Shuffle/Sort
- Generan la salida final

Map Reduce

CONTAR PALABRAS

En un lugar de la Mancha
El lugar era un desastre
No hay ropa sin mancha

MAP

REDUCE

RESULTADOS

De	1
El	1
En	1
Era	1
Hay	1
La	1
Lugar	2
Mancha	2
No	1
Ropa	1
Sin	1
Un	2

Map Reduce

CONTAR PALABRAS

En un lugar de la Mancha
El lugar era un desastre
No hay ropa sin mancha



En	1
Un	1
Lugar	1
De	1
La	1
mancha	1



El	1
Lugar	1
Era	1
Un	1
desastre	1



No	1
hay	1
Ropa	1
Sin	1
mancha	1

Map Reduce

CONTAR PALABRAS

En un lugar de la Mancha
El lugar era un desastre
No hay ropa sin mancha



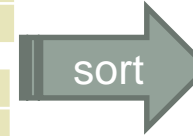
En	1
Un	1
Lugar	1
De	1
La	1
mancha	1



El	1
Lugar	1
Era	1
Un	1
desastre	1



No	1
hay	1
Ropa	1
Sin	1
mancha	1



De	1
El	1
En	1
Era	1
Hay	1
La	1
Lugar	1,1
Mancha	1,1
No	1
Ropa	1
Sin	1
Un	1,1

Map Reduce

