

ARQUITECTURA DE SOFTWARE



**UNIVERSIDAD
DE ANTIOQUIA**
1 8 0 3

DOCENTE

Robinson Coronado

Bryan Zuleta
Weimar Quintero

Facultad de Ingeniería
Ingeniería de Sistemas
Universidad de Antioquia
2019-2

El Map Reduce es un modulo del framework **APACHE HADOOP** que permite el procesamiento de grandes conjuntos de datos en varios ordenadores, pero que a ojos del usuario parece un único ordenador. Hadoop separa y distribuye automáticamente los archivos que contienen los datos, además de dividir el trabajo que tiene en tareas muchos más pequeñas y ejecutarlas de manera distribuida y recuperarse de posibles fallos automáticamente y de una forma transparente al usuario.

MAP REDUCE PATTERN

¿Qué Es?

Es un modelo de programación para dar soporte a la computación paralela sobre grandes colecciones de datos de grupos de computadoras. MapReduce se emplea en la resolución práctica de algunos algoritmos susceptibles a ser paralelizados. No obstante MapReduce no es la solución para cualquier problema, de la misma forma que cualquier problema no puede ser resuelto eficientemente por MapReduce. Por lo general se abordan problemas de dataset de grandes tamaños, alcanzando hasta los petabytes de tamaño. Es por esta razón que este framework suele ejecutarse en sistemas de archivos distribuidos. No todos los procesos pueden ser abordados por MapReduce, concretamente son abordables solo aquellos que se pueden disgregar en dos operaciones las cuales son: Operación MAP y Operación REDUCE, porque ambas están definidas con respecto a datos estructurados.

MapReduce maneja dos etapas, la etapa mapa y la etapa reducir

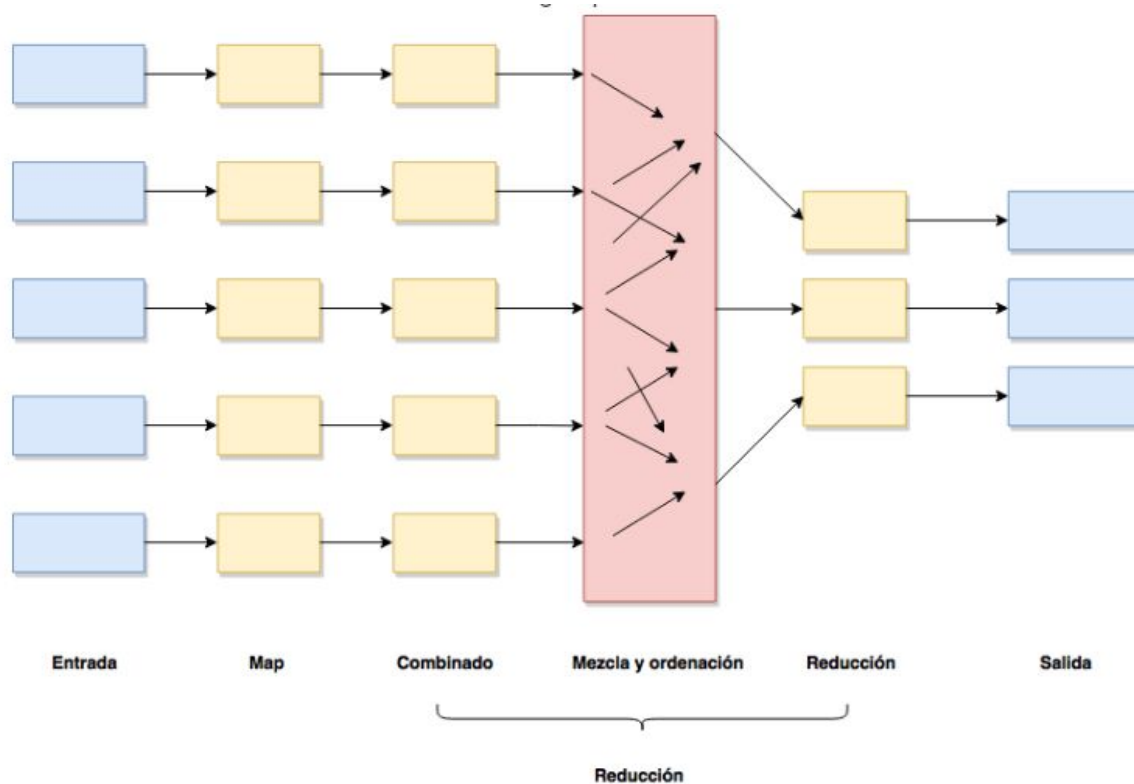
Etapas de Mapeo: Esta etapa consiste en procesar los datos de entrada, por lo general los datos de entrada se encuentran en una forma de archivo o directorio y se almacena en el sistema de archivos HADOOP. El archivo de entrada se pasa en la función de mapa línea por línea, y luego procesa los datos y crea varios pequeños fragmentos de datos.

Etapas de Reducción: Esta etapa es la combinación de la reproducción aleatoria y la etapa de reducción, el trabajo de esta es la de procesar los datos que llegan desde el mapa, después de un proceso de elaboración, se genera un nuevo conjunto de producción, que se almacena en el HDFS.

El HDFS es el distribuidor de sistema de archivos de Hadoop , la cual su objetivo es la capacidad para almacenar los archivos en un cluster de varias máquinas.

Esta característica es dominante cuando se pretenden almacenar grandes cantidades de datos, puesto que en general no es posible almacenar cientos de terabytes o petabytes en una única máquina.

DIAGRAMA DE COMO FUNCIONA EL MAP REDUCE.



Y COMO SE APLICA EN LA VIDA REAL

Un ejemplo simple como lo voy a poner aquí sería manejar el salario de un personal, es decir, quiero saber el salario total de todo el personal basado en el genero.

Scenario	Counting the total salary of employees based on gender.
Map (Key, Value)	Key : Gender Value : Their Salary
Reduce	Group by Key (Gender), then take the sum of the value (salary) for each group.

Ahora la cantidad de personas y su salario basado en miles

EMPLOYEE NAME	GENDER	SALARY IN K
Krishna	Male	80
Namrata	Female	75
Ritesh	Male	100
Ram	Male	40
Vijay	Male	250
Jitu	Male	55
Ria	Female	90
Poonam	Female	300
Radha	Female	40

Luego de eso va a pasar lo siguiente

To count the total salary by gender, we need to make the key Gender and the value Salary. The output for the Map function is:

```
(Male, 80), (Female, 75), (Male, 100), (Male, 40), (Male, 250),
(Male, 55), (Female, 90), (Female, 300), (Female, 40)
```

Intermediate splitting gives the input for the Reduce function:

```
(Male <80, 100, 40, 250, 55>), (Female<75, 90, 300, 40>)
```

And the Reduce function output is:

```
(Male, 525), (Female, 505)
```

Primero el va a tomar cada persona y lo va a encerrar en un mapa, en donde cada hombre y mujer que manejan un valor, los guarda en una llave, luego de eso, agrupa en otra llave, todos los salarios de todos los hombres y luego el de todas las mujeres, y al final reduce aún más la función haciendo que sume todos los salarios de los hombres y mujeres y los deje en una llave con el género y el total.

WEBGRAFÍA

https://www.tutorialspoint.com/es/hadoop/hadoop_mapreduce.htm

<https://dzone.com/articles/mapreduce-design-patterns-1>

<https://geekytheory.com/fundamentos-de-apache-hadoop-y-mapreduce>