
MapReduce

- Ortega Ibarra Jaime Jesús.
 - Martiñón Luna Jonathan José.
 - José de Jesús Tapia López.
-

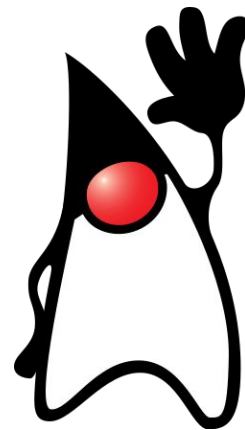
¿Qué es?

MAPREDUCE es un modelo de programación para dar soporte a la computación paralela sobre grandes colecciones de datos en grupos de computadoras.

El nombre del framework está inspirado en los nombres de dos importantes métodos de la programación funcional:

Map y Reduce.

Se han escrito implementaciones de bibliotecas de MapReduce en diversos lenguajes de programación como C++, Java y Python.





MAP()

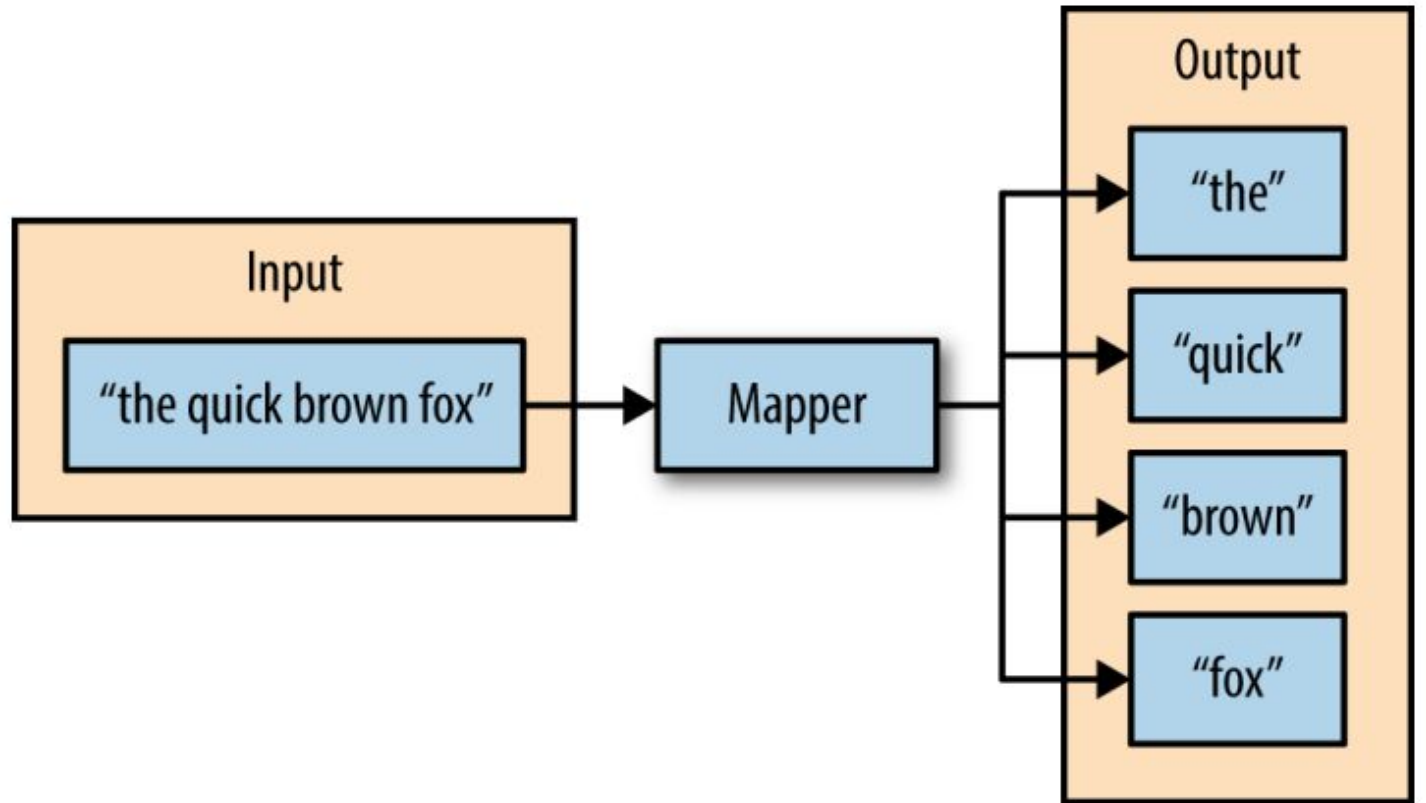
Las funciones Map y Reduce están definidas ambas con respecto a datos estructurados en tuplas del tipo (clave, valor).

- Map toma uno de estos pares de datos con un tipo en un dominio de datos, y devuelve una lista de pares en un dominio diferente.

```
Map( $k_1, v_1$ ) -> list( $k_2, v_2$ )
```

- Se encarga del mapeo y es aplicada en paralelo para cada ítem en la entrada de datos. Esto produce una lista de pares (k_2, v_2) por cada llamada.

Ejemplo MAP:





Reduce()

- La función reduce es aplicada en paralelo para cada grupo, produciendo una colección de valores para cada dominio.

```
Reduce( $k_2$ , list ( $v_2$ )) -> list( $v_3$ )
```

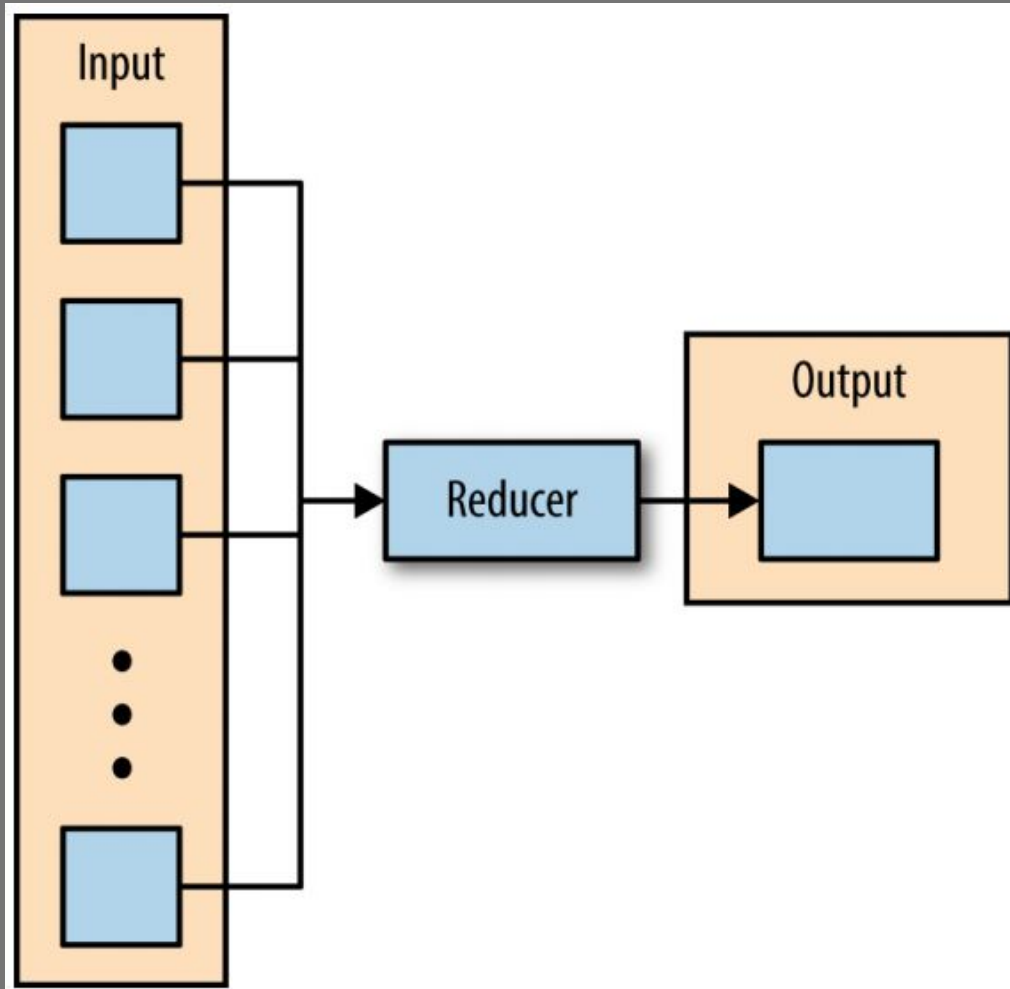
- Para cada llamada a Reduce típicamente produce un valor v_3 o una llamada vacía, aunque una llamada puede retornar más de un valor. El retorno de todas esas llamadas se recoge como la lista de resultado deseado.



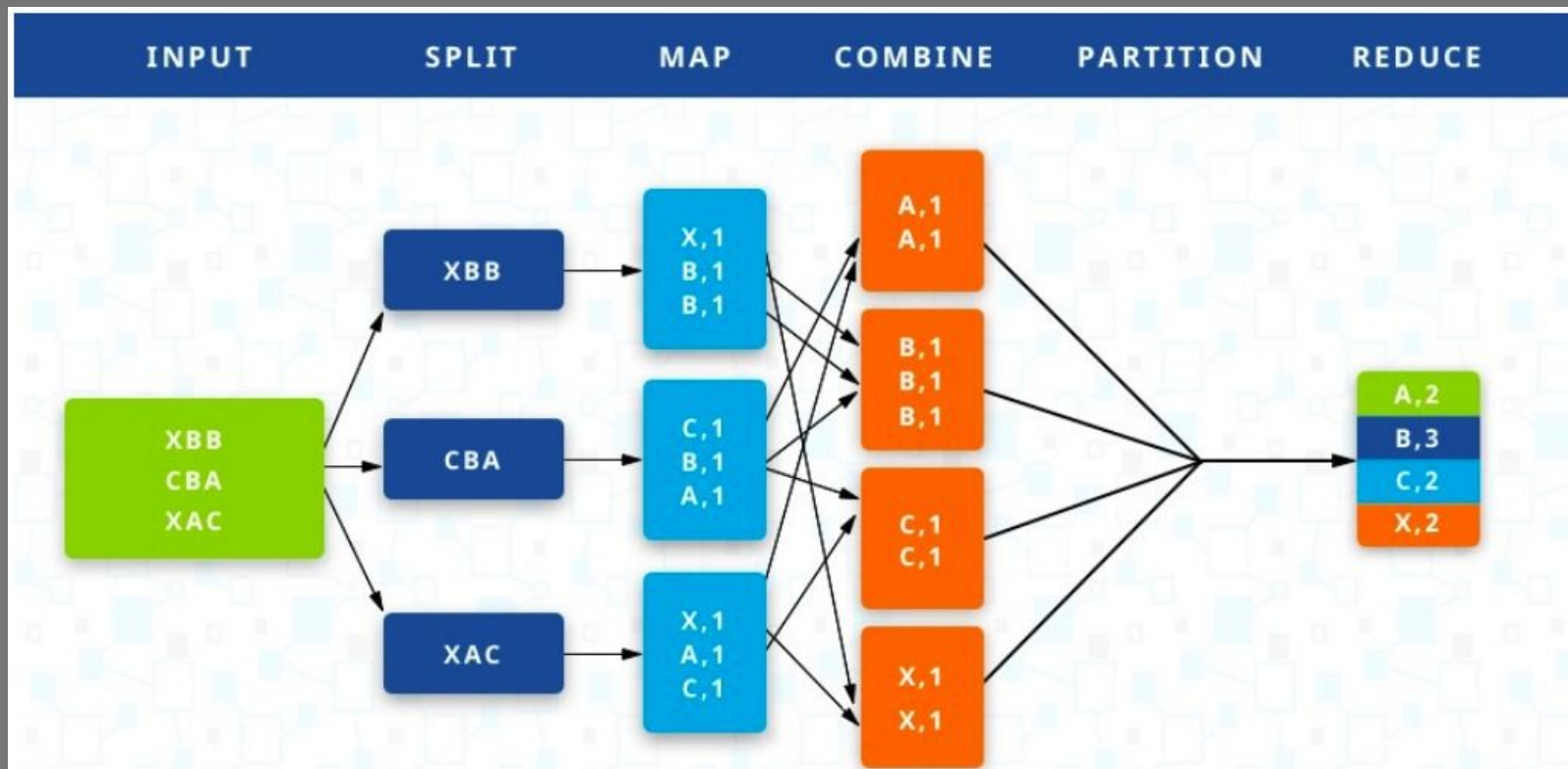
Por lo tanto

El framework MapReduce transforma una lista de pares (clave, valor) en una lista de valores. Este comportamiento es diferente de la combinación "map and reduce" de programación funcional, que acepta una lista arbitraria de valores y devuelve un valor único que combina todos los valores devueltos por map.

Ejemplo REDUCE:



Ejemplo Map - Reduce:



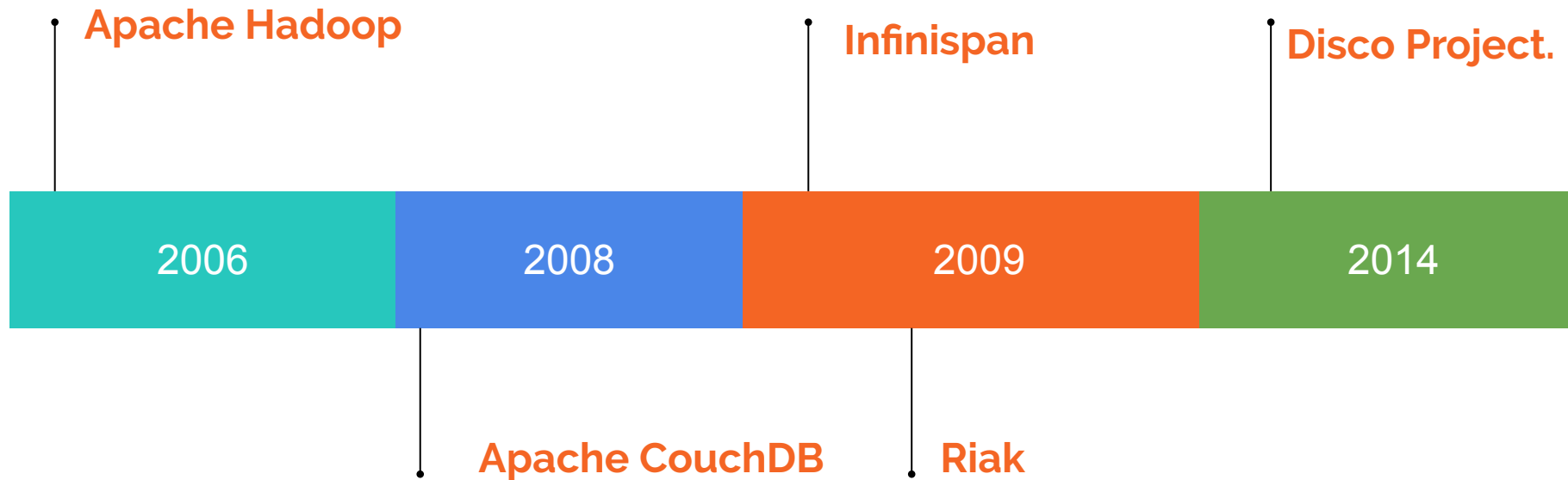
Ventajas.

- Automatiza aspectos de paralelismo y tolerancia a fallos.
- Permite que programador se ocupe sólo de funcionalidad requerida.
- Mejora el código al distribuir las tareas y simplificar.
- Genera beneficios en cuestión de tiempo al poder ser paralelo.

Desventaja

- Al ser paralelo no siempre va a ser eficiente, esto va a depender del problema que se desea solucionar.
 - Su implementación puede llegar a ser muy compleja.
-

Implementaciones.



Ejemplo:

Aquí pondría mi ejemplo...



Si hubiera encontrado uno!!!

Referencias:

- Radtka, Z. & Miner, D. (2015). Hadoop with Python. United States of America: O'Reilly.
 - Olmedo, Y. (2012). ¿Qué es MapReduce?. noviembre 21, 2019, de SolidQ Sitio web: <https://blogs.solidq.com/es/business-analytics/que-es-mapreduce/>
 - Robert R.F. (2018). Your First Map Reduce—Using Hadoop with Python and OSX. noviembre 21, 2019, de Medium Sitio web: <https://medium.com/@rrfd/your-first-map-reduce-using-hadoop-with-python-and-osx-ca3b6f3dfe78>
-