

## Proyecto de programación avanzada

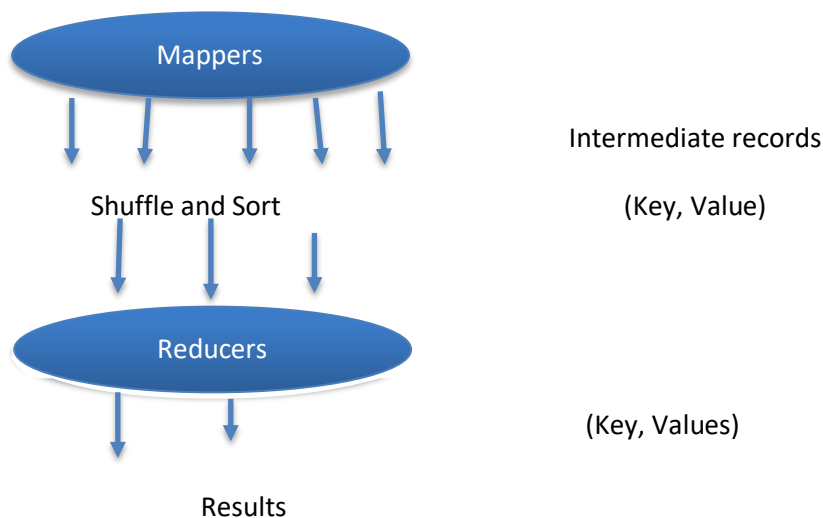
En el mundo de “big data” se ha venido utilizando el modelo de programación MapReduce compuesto de un procedimiento Map que realiza las funciones de ordenamiento, filtrado, etc. y un procedimiento de reducción que realiza una operación de acumulado o agregación de lo producido por los procesos “mapeadores”.

El procesar un gran archivo de manera secuencial de principio a fin toma demasiado tiempo, ¿Que pasaría si partiéramos el archivo en segmentos y procesáramos cada archivo en paralelo? Seguramente, el proceso se haría más rápido.

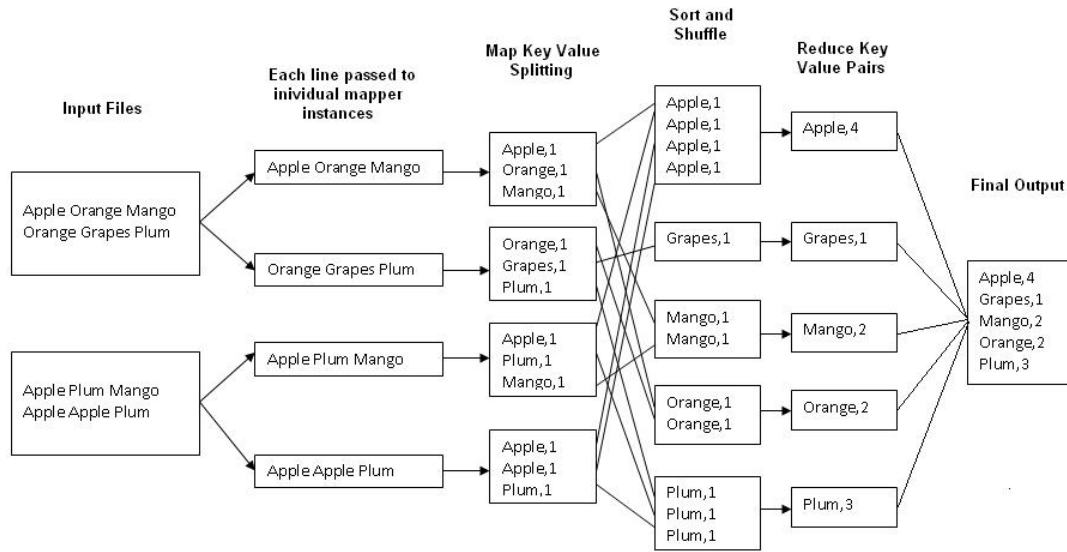
Los procesos “Mapeadores” clasifican la entrada en colecciones llave-valor:

- Llave lo que es importante (de acuerdo con la consulta)
- Valor que tan importante es

Los procesos Reductores toman la salida de los “Mapeadores” y realizan operaciones de agregación



## Proyecto de programación avanzada



Su labor es crear varios procesos “mapeadores” (al menos 3) y un reductor para realizar el conteo de palabras.

Como parte de la entrega, deberán escribir un artículo de divulgación científica que utilice el formato de la IEEE (archivo adjunto en Canvas) con las siguientes secciones:

- 1) Autores y filiaciones
- 2) Resumen
- 3) Introducción
- 4) Estado del arte. Incluyan descripciones de lo que es el modelo de programación MapReduce sus orígenes en los lenguajes funcionales. Un artículo a revisar es: <https://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf>
- 5) Desarrollo. Explique los experimentos realizados, como lo realizaron, como el lector de su artículo puede reproducir el experimento, etc.
- 6) Resultados obtenidos
- 7) Conclusiones y trabajo futuro
- 8) Bibliografía. Es muy importante citar a todo lo que le da soporte a su discurso. No pongan una lista de referencias de las cuales no se encuentren citadas en su discurso

Las ponderaciones es de 50% producto funcionando 50% calidad del artículo entregado y presentación.