

## MapReduce - Ejercicios

### Ejercicio 1

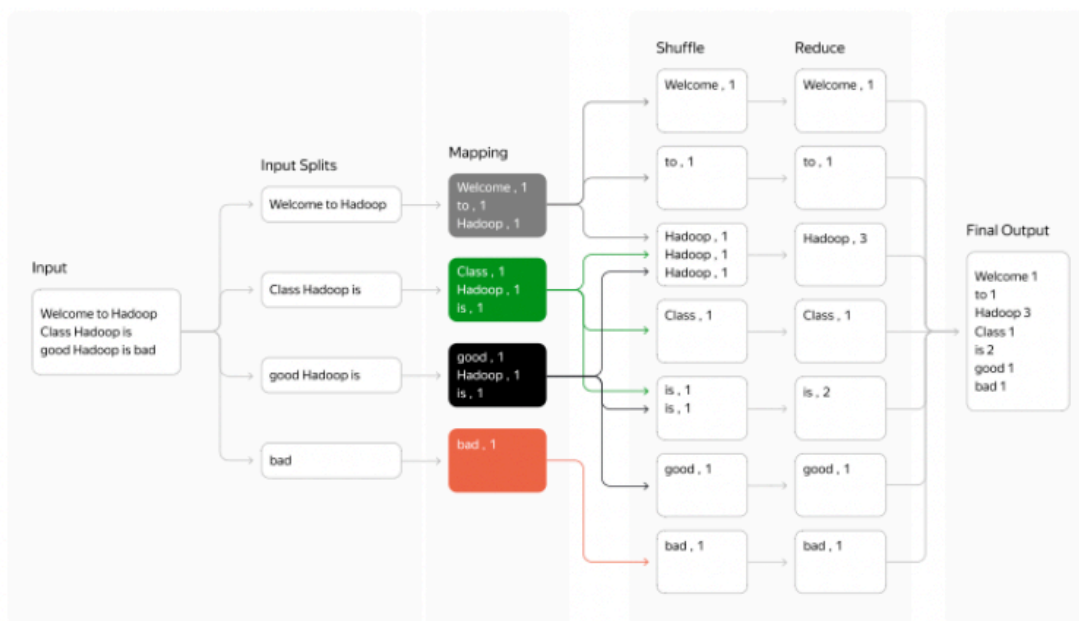
Una empresa desea analizar la frecuencia de palabras en un conjunto de datos de texto utilizando el modelo de procesamiento distribuido MapReduce. Se cuenta con **4 máquinas** para realizar el procesamiento distribuido.

#### Input

“Las empresas españolas suelen ser empresas pequeñas y medianas empresas”.

- Divide el texto en los splits necesarios y simula la asignación a máquinas.
- Indica qué hará cada etapa del modelo MapReduce hasta obtener la salida final.
- Resuelve el ejercicio en Python, dividiendo el texto en fragmentos, realizando las etapas de Mapping, Shuffle y Reduce, y mostrando los resultados finales.

Para resolver los apartados a) y b) sigue este ejemplo:



### Ejercicio 2

Una empresa de logística desea analizar el tiempo de entrega de paquetes en diferentes ciudades utilizando el modelo de procesamiento distribuido MapReduce, para identificar aquellas con mayores retrasos. El objetivo es calcular el promedio de retraso en días por ciudad a partir de un conjunto de datos distribuidos. Para ello, se cuenta con **3 máquinas**. Los registros se almacenan de la forma ["Ciudad, Retraso"].

A partir de estos datos:

["Madrid, 2", "Barcelona, 1", "Madrid, 3", "Valencia, 0", "Barcelona, 2", "Madrid, 1"]

- a) Divide el texto en los splits necesarios y simula la asignación a máquinas.
- b) Indica qué hará cada etapa del modelo MapReduce hasta obtener la salida final.
- c) Resuelve el ejercicio en Python, dividiendo el texto en fragmentos, realizando las etapas de Mapping, Shuffle y Reduce, y mostrando los resultados finales.

Resuelve el ejercicio utilizando el modelo de procesamiento distribuido MapReduce. Para resolver los apartados a) y b) sigue el ejemplo mostrado en el ejercicio 1.

### Ejercicio 3

---

Una empresa de ventas online desea analizar el rendimiento de sus productos durante un periodo de tiempo utilizando el modelo de procesamiento distribuido MapReduce. Cada registro contiene el ID del producto y el ingreso generado por su venta, en formato: ["ProductoID, Ingreso"].

Se cuenta con los siguientes datos:

["P001, 200", "P002, 450", "P001, 300", "P003, 150", "P002, 550", "P001, 250", "P003, 350", "P002, 400"].

El objetivo es calcular:

1. La **suma total de ingresos** por producto.
2. El **ingreso promedio** por producto.
3. El **ingreso máximo y mínimo** obtenido por producto.

Cuentan con 3 máquinas para realizar el proceso distribuido.

- a) Divide el texto en los splits necesarios y simula la asignación a máquinas.
- b) Indica qué hará cada etapa del modelo MapReduce hasta obtener la salida final.
- c) Resuelve el ejercicio en Python, dividiendo el texto en fragmentos, realizando las etapas de Mapping, Shuffle y Reduce, y mostrando los resultados finales.

Para resolver los apartados a) y b) sigue el ejemplo mostrado en el ejercicio 1.

### Ejercicio 4

---

Un analista de sistemas ha obtenido este esquema al aplicar MapReduce a un problema que le ha planteado su empresa. Indica el enunciado inicial del que ha partido y crea un programa en Python que lo resuelva.

