



Universidad de Córdoba

MÁSTE EN INTELIGENCIA COMPUTACIONAL E INTERNET DE LAS COSAS

# EJERCICIOS PROPUESTOS: EL PARADIGMA MAP-REDUCE

Autor: Alberto Fernández Merchán  
Asignatura: Introducción al Big Data Analytics

## 1. Ejercicio 1

1. Para aquellas solicitudes a páginas php respondidas exitosamente, queremos saber cuántos accesos únicos (distintos clientes) ha tenido cada página.

- **Entradas y salidas:**

- **Entrada:** Un conjunto de registros con columnas IP, URL y Status.
- **Salida:** Un par (URL, conteo de IPs únicas) donde la URL corresponde a páginas .php con respuestas exitosas.

- **Funciones map() y reduce():**

- **map():** Filtra las filas relevantes y emite pares (URL, IP).
- **reduce():** Agrupa por URL y cuenta las direcciones IP únicas.

- **Secuencia completa:**

- a) Filtrar registros con URLs que contienen “.php” y códigos de estado 2xx.
- b) Aplicar map() para generar pares (URL, IP).
- c) Reducir por clave (URL) y contar las IPs únicas.

2. Queremos conocer la frecuencia de acceso de cada cliente a recursos de nuestro servidor.

- **Entradas y salidas:**

- **Entrada:** Un conjunto de registros con columnas IP, URL y Status.
- **Salida:** Un par (IP, conteo de accesos) donde el conteo es el número total de accesos por cliente.

- **Funciones map() y reduce():**

- **map():** Emite pares (IP, 1) para cada registro.
- **reduce():** Suma los valores por clave (IP) para obtener el conteo total de accesos.

- **Secuencia completa:**

- a) Aplicar map() para generar pares (IP, 1).
- b) Reducir por clave (IP) sumando los valores asociados.

3. Con el fin de detectar posibles ataques o errores, queremos conocer el número de respuestas no correctas que se han devuelto a los clientes que han recibido al menos un error.

- **Entradas y salidas:**

- **Entrada:** Un conjunto de registros con columnas IP y Status.
- **Salida:** Un par (IP, conteo de errores) donde el conteo representa el número de respuestas no correctas.

- **Funciones map() y reduce():**

- **map():** Filtra filas con códigos de estado entre 400 y 599 y emite pares (IP, 1).
- **reduce():** Suma los valores por clave (IP) para obtener el conteo total de errores.

- **Secuencia completa:**

- a) Filtrar registros con códigos de estado entre 400 y 599.
- b) Aplicar map() para generar pares (IP, 1).
- c) Reducir por clave (IP) sumando los valores asociados.

## 2. Ejercicio 2

URL de los ejercicios:

- Ejercicio 2: [https://colab.research.google.com/drive/1CeGY2Zu-NgeCA8d\\_Grxbvd90JbizwCnB?usp=drive\\_link](https://colab.research.google.com/drive/1CeGY2Zu-NgeCA8d_Grxbvd90JbizwCnB?usp=drive_link)
- Ejercicio 4: [https://colab.research.google.com/drive/1wjPgC23f7aYKApMKAsjWMRKxaQ9NZOLa?usp=drive\\_link](https://colab.research.google.com/drive/1wjPgC23f7aYKApMKAsjWMRKxaQ9NZOLa?usp=drive_link)
- Ejercicio 5: [https://colab.research.google.com/drive/113ghWL7d9k0\\_GuHwVsDV1RDqB79g7hyj?usp=drive\\_link](https://colab.research.google.com/drive/113ghWL7d9k0_GuHwVsDV1RDqB79g7hyj?usp=drive_link)