# Índice General

Tarea Bloque 1 MapReduce 3

Ejercicio B 3

**Esquema** 3

**Explicación** 4

**Pseudocódigo** 4

Ejercicio D 4

**Esquema** 5

**Explicación** 5

**Pseudocódigo** 5

# Índice de Figuras

[**Figura 1:** Esquema MapReduce ejercicio B 3](#_Toc188235442)

[**Figura 2:** Esquema MapReduce ejercicio D 5](#_Toc188235443)

# MapReduce

## Ejercicio B

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

### **Esquema**

Diagrama

Descripción generada automáticamente

**Figura 1:** Esquema MapReduce ejercicio B

### **Explicación**

Inicialmente es necesario realizar un filtrado, se seleccionan únicamente las entradas correspondientes a páginas .php con respuestas exitosas. Posteriormente, en la fase Map, se genera un par clave-valor donde la clave es la URL de la página y el valor representa un acceso. Luego, en la etapa Shuffle, los pares con la misma clave se agrupan, y finalmente, la fase Reduce suma los valores asociados a cada clave para obtener el número total de accesos únicos por página.

### **Pseudocódigo**

#### Función de Filtrado

**Filtrar (**URL**,** status**,** IP**):**

**return** **(**URL**,** IP**)** contiene ".php" **and** status **>=** 200 **and** status **<=** 299

#### Función Map

**map** **(**key**,** values**):**

**//** keys**:** weblog**,** values**:** URLs

**for** each URL u **in** values**:**

Emit **(**URL**,** "1"**);**

#### Función Reduce

**reduce** **(**key**,** lista\_de\_valores**):**

**//** keys**:** URLs**,** valores**:** lista **del** conteo

suma **=** 0

**for** valor **in** lista\_de\_valores**:**

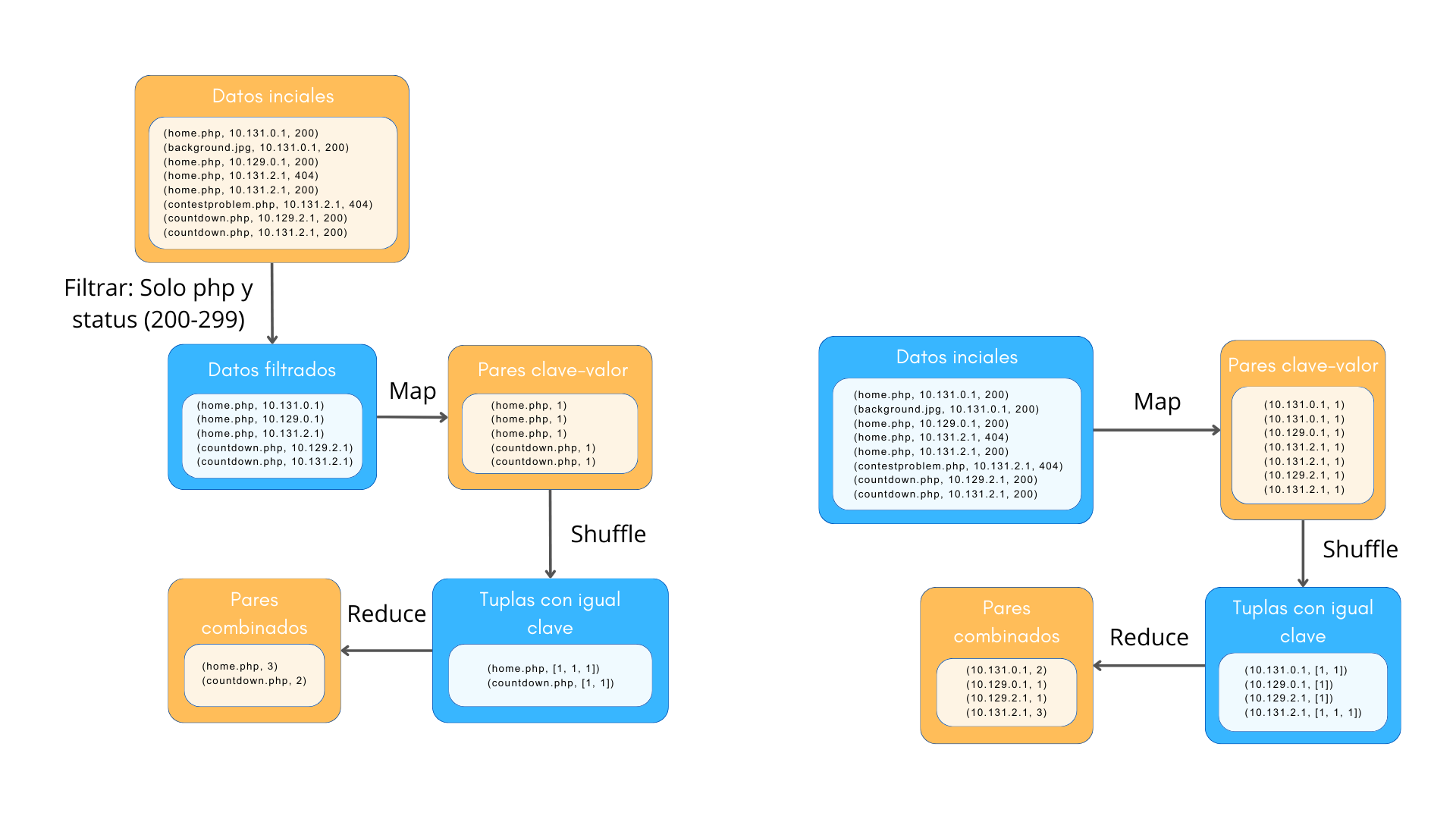
suma **+=** valor

Emitir**(**key**,** suma**)**

## Ejercicio D

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

### **Esquema**



**Figura 2:** Esquema MapReduce ejercicio D

### **Explicación**

En este caso, no es necesario aplicar un filtrado previo. Directamente en la fase Map, cada registro genera un par clave-valor donde la clave es la dirección IP del cliente y el valor es un contador (1). En la etapa Shuffle, se agrupan las claves idénticas (las mismas IPs), y en la fase Reduce, se suman los valores asociados para calcular la cantidad total de accesos realizados por cada cliente, representando la frecuencia de uso.

### **Pseudocódigo**

#### Función Map

**map** **(**key**,** values**):**

**//** keys**:** weblog**,** values**:** IPs

**for** each IP **in** values**:**

Emit **(**IP**,** "1"**);**

#### Función Reduce

**reduce** **(**key**,** lista\_de\_valores**):**

**//** keys**:** IPs**,** valores**:** lista **del** conteo

suma **=** 0

**for** valor **in** lista\_de\_valores**:**

suma **+=** valor

Emitir**(**key**,** suma**)**