Modelo de procesamiento MapReduce

Problema: Calcular la temperatura máxima de cada año

□ Subproblemas: Por cada año, calcular la temperatura máxima

Clave: Año Valor: Temperatura

- Problema: Calcular la temperatura máxima de cada año
 - □ Subproblemas: Por cada año, calcular la temperatura máxima
 - Clave: Año
 Valor: Temperatura

```
1999, 2°, Lunes...

2000, 3°, Miércoles...

2000, 1°, Martes,...

1999, 5°, Domingo...

1999, 7°, Sábado...

1999, 3°, Lunes...

2000, 4°, Martes...

2000, 4°, Martes...

2000, 1°, Sábado...

1999, 6°, Lunes...

Map <1999, 2°>

<1999, [2°, 5°,

7°, 3°, 6°]>

Reduce <1999, 7°>

Map <1999, 3°>

Reduce <2000, 4°>

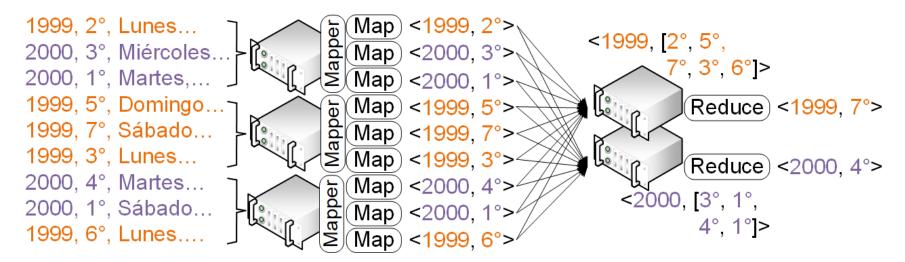
Map <2000, 4°>

Map <2000, 1°>

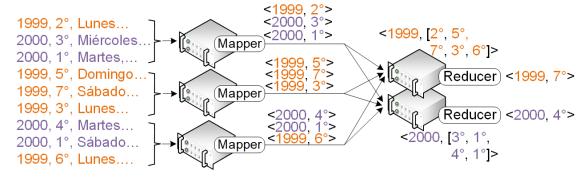
Map <
```

- Problema: Calcular la temperatura máxima de cada año
 - Subproblemas: Por cada año, calcular la temperatura máxima

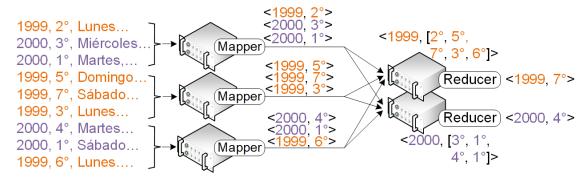
Clave: AñoValor: Temperatura

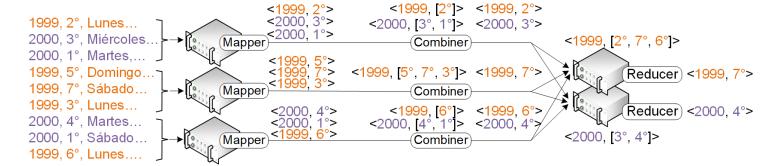


Problema: Calcular la temperatura máxima de cada año



Problema: Calcular la temperatura máxima de cada año





Modelo de procesamiento MapReduce

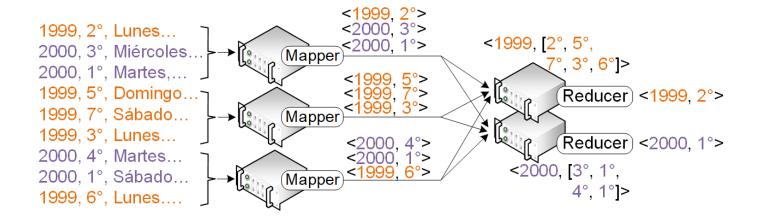
Problema: Calcular la temperatura mínima de cada año

□ Subproblemas: Por cada año, calcular la temperatura mínima

Clave: AñoValor: Temperatura

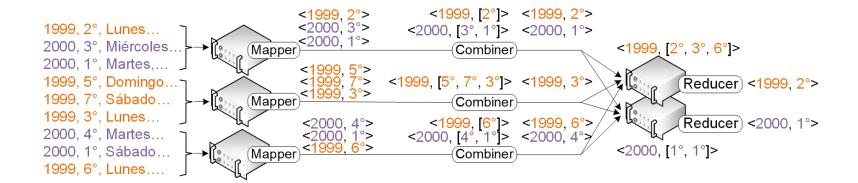
- Problema: Calcular la temperatura mínima de cada año
 - Subproblemas: Por cada año, calcular la temperatura mínima

Clave: AñoValor: Temperatura

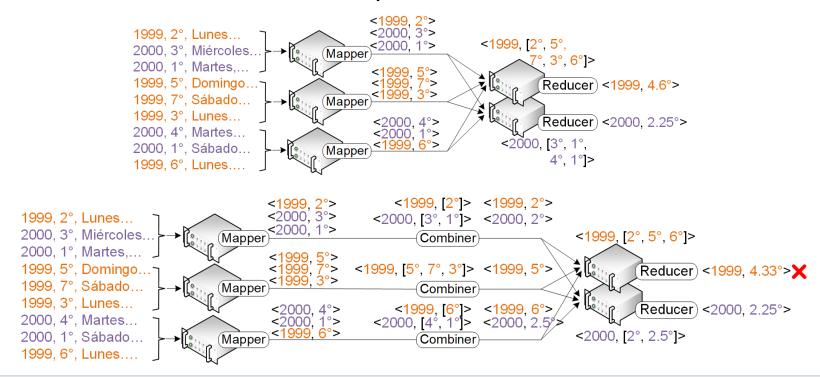


- Problema: Calcular la temperatura mínima de cada año
 - Subproblemas: Por cada año, calcular la temperatura mínima

Clave: Año Valor: Temperatura

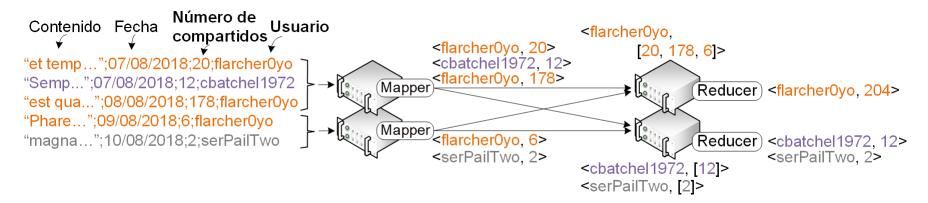


Problema: Calcular la temperatura media de cada año

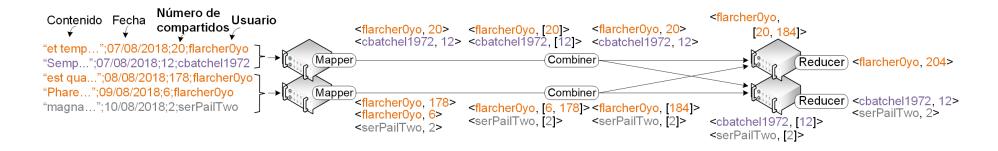


> 25

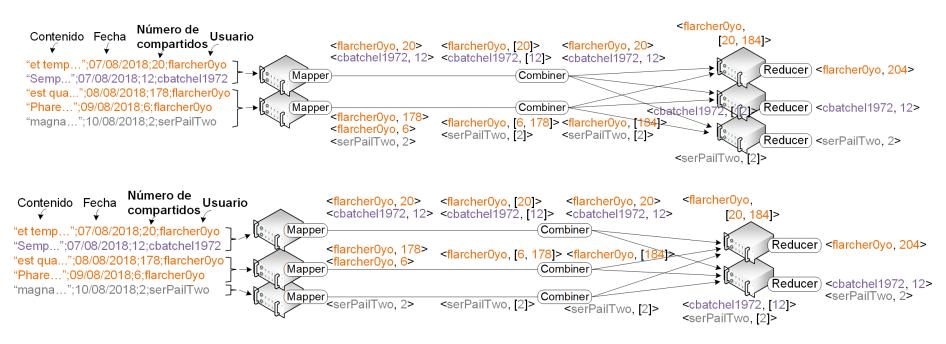
- Problema: Obtener el número de veces que se compartieron los tweets de cada usuario
 - Subproblemas: Por cada usuario, obtener el número de veces que se compartieron sus tweets
 - Clave: Usuario
 Valor: número de veces que se comparten sus tweets



- Problema: Obtener el número de veces que se compartieron los tweets de cada usuario
 - □ Subproblemas: Por cada usuario, obtener el número de veces que se compartieron sus tweets
 - Clave: Usuario
 Valor: número de veces que se comparten sus tweets



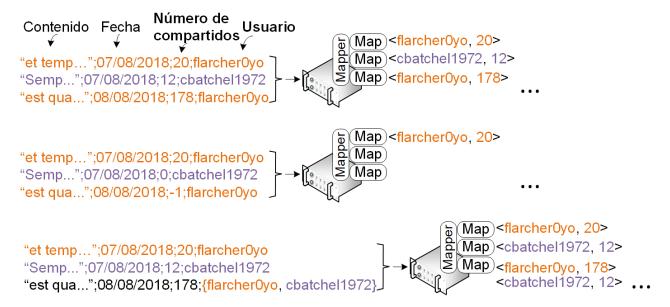
 Problema: Obtener el número de veces que se compartieron los tweets de cada usuario



> 28

Viu Universidad Internacional de Valencia

- Mapper: Ejecuta Map por cada dato de entrada
 - □ Map puede emitir {0, 1, n} pares <clave, valor>



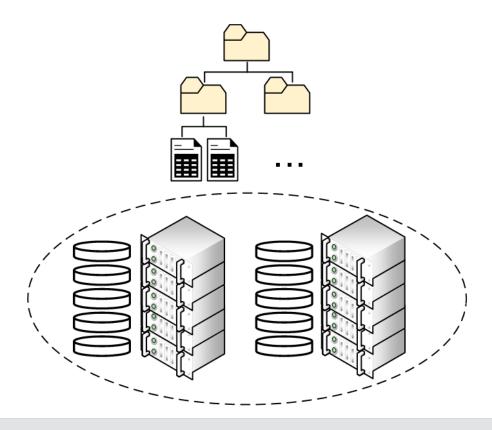
- Mapper: Ejecuta Map por cada dato de entrada
 - □ Map puede emitir {0, 1, n} pares <clave, valor>
- Reducer: Ejecuta Reduce por cada "subproblema"
 - □ Reduce puede emitir {0, 1, n} pares <clave, valor>
 - Seleccionar los usuarios que tienen más de 1000 compartidos
 - Seleccionar por cada usuario el número de tweets publicados
 - Seleccionar por cada autor los 10 tweets con más compartidos

- Sencillo
- Distribuido
- Escalable
- Big Data

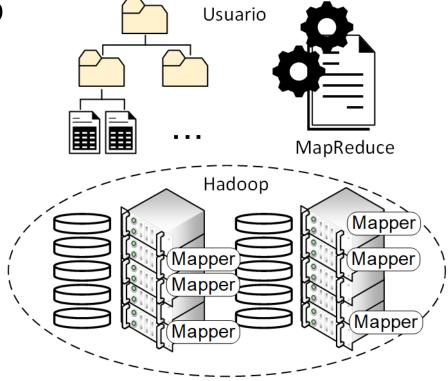
- Frameworks/implementaciones MapReduce:
 - ☐ Google MapReduce
 - □ Hadoop MapReduce
 - □ Spark
 - □ Flink
 - □ MongoDB
 - □ Riak
 - □ ...

- Hadoop:
 - □ Popular
 - Maduro
 - Extendible
 - □ Open Source

HDFS



Hadoop



- Hadoop
 - ☐ Sistema de archivos: HDFS, S3,...
 - □ Motor de procesamiento: Hadoop MapReduce,...
 - □ Core
 - Ecosistema