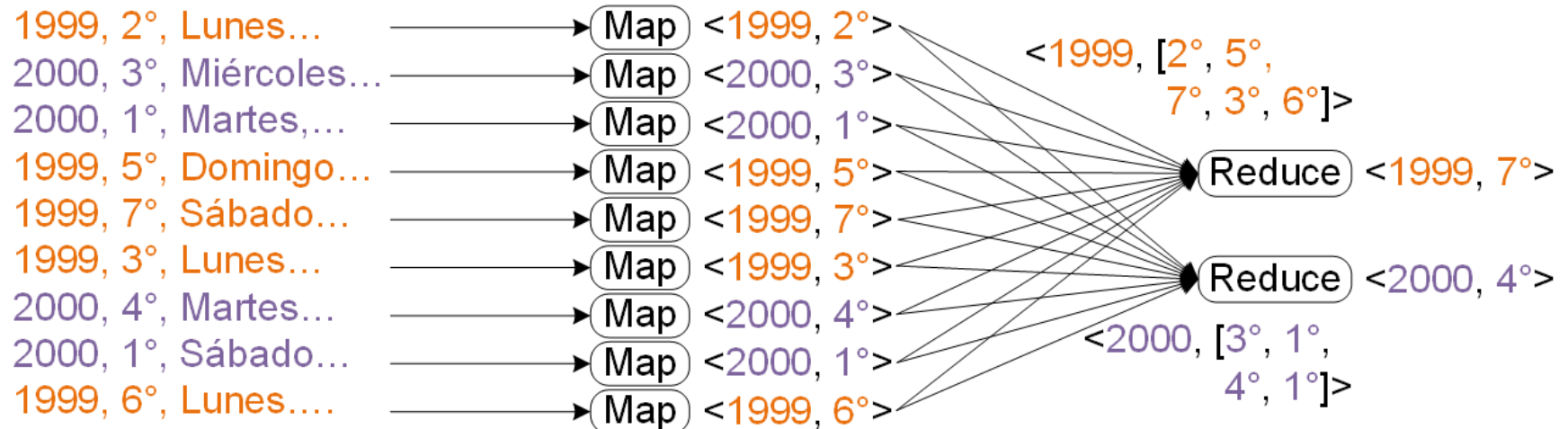


- Problema: Calcular la temperatura máxima de cada año
 - Subproblemas: Por cada año, calcular la temperatura máxima
 - Clave: Año Valor: Temperatura

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

■ Clave: Año

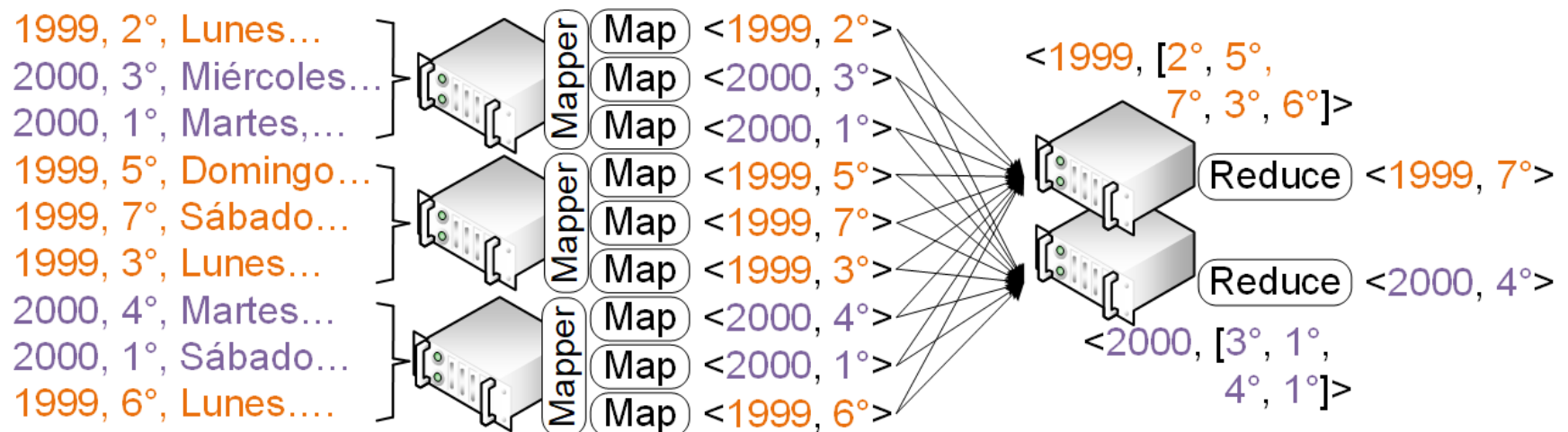
Valor: Temperatura



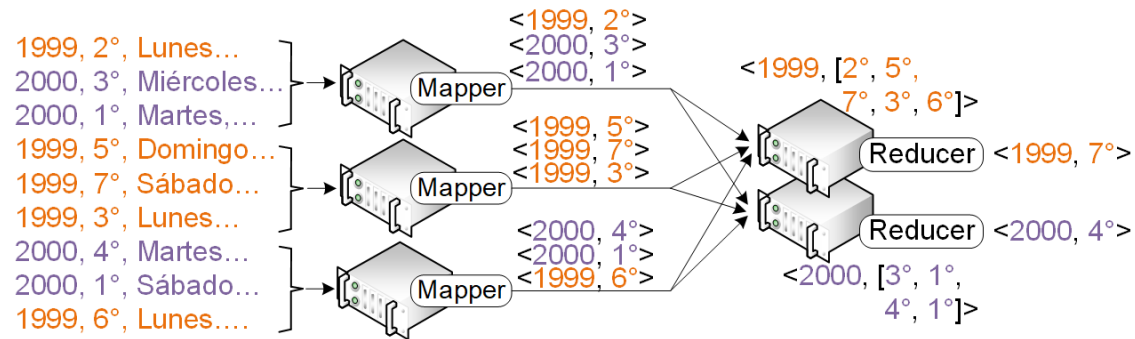
- Problema: Calcular la temperatura máxima de cada año
 - Subproblemas: Por cada año, calcular la temperatura máxima

■ Clave: Año

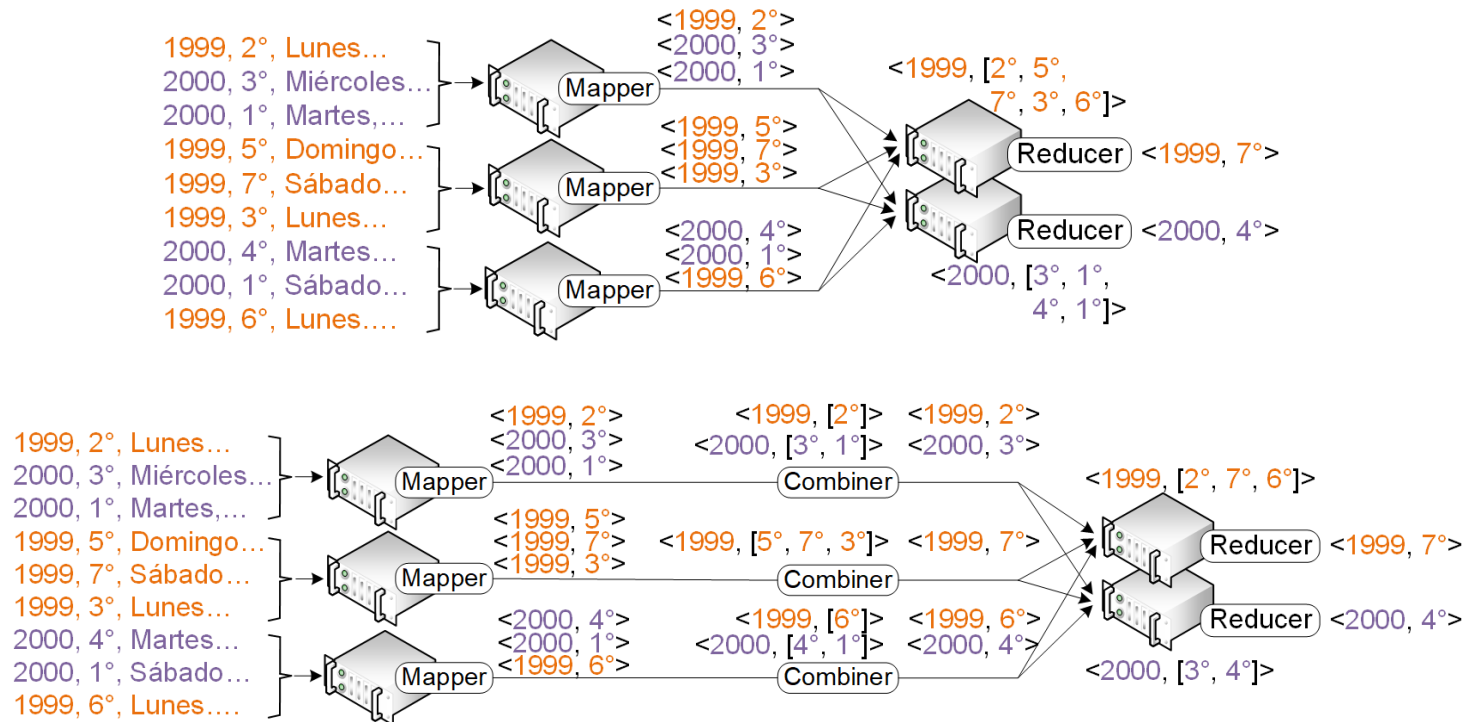
Valor: Temperatura



- Problema: Calcular la temperatura máxima de cada año



■ Problema: Calcular la temperatura máxima de cada año

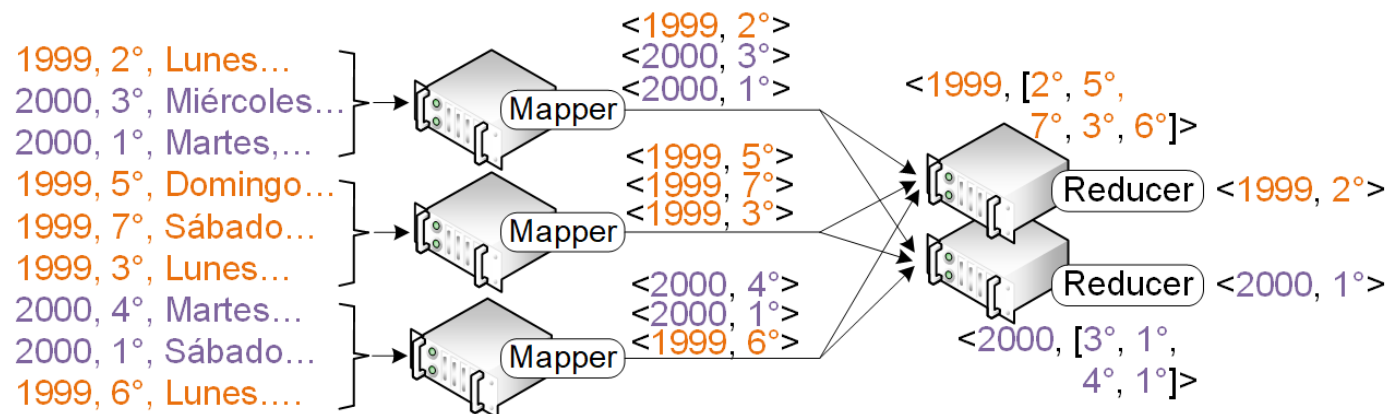


- Problema: Calcular la temperatura mínima de cada año
 - Subproblemas: Por cada año, calcular la temperatura mínima
 - Clave: Año Valor: Temperatura

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

■ Clave: Año

Valor: Temperatura

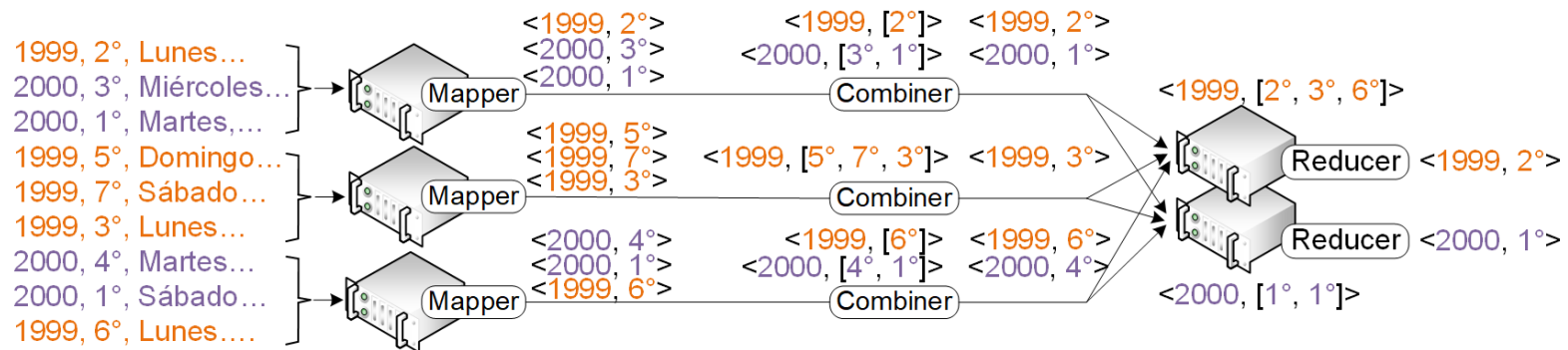


- Problema: Calcular la temperatura mínima de cada año

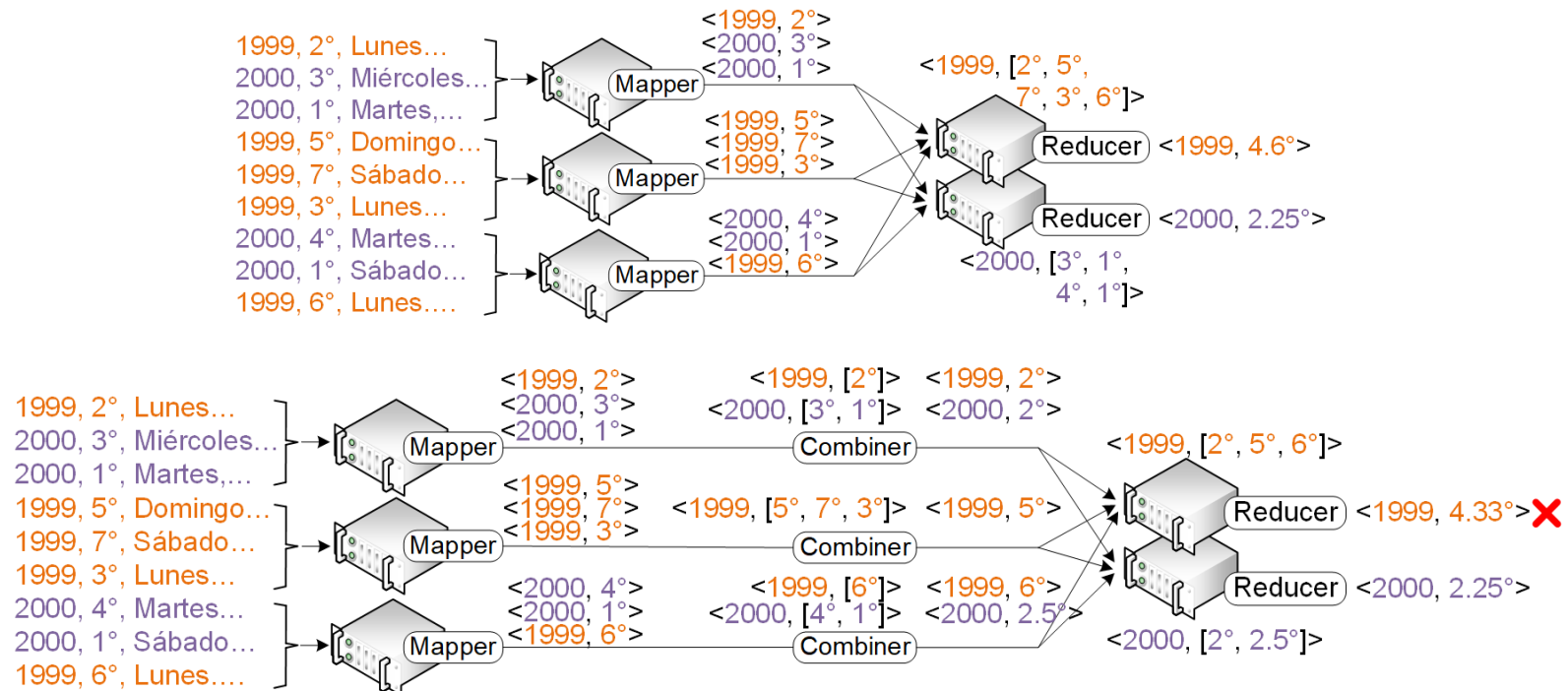
- Subproblemas: Por cada año, calcular la temperatura mínima

- Clave: Año

- Valor: Temperatura

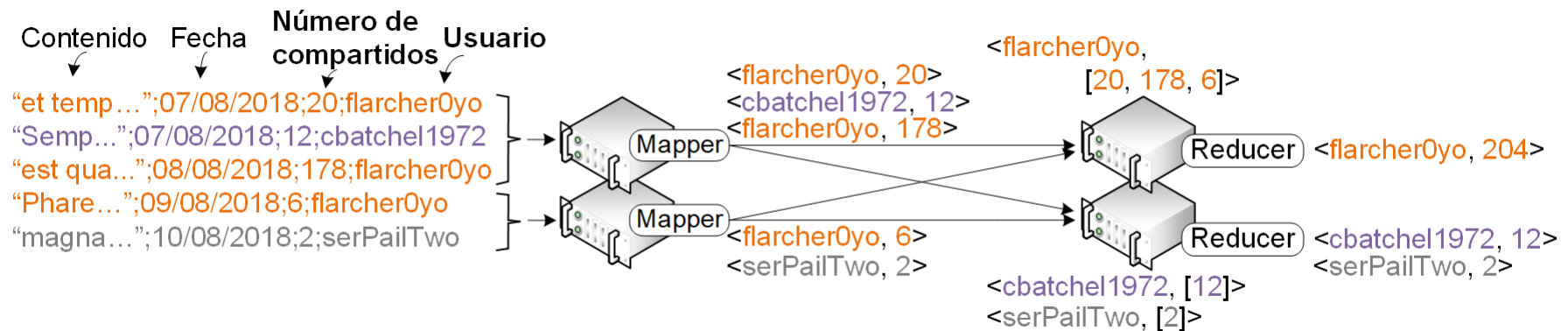


- Problema: Calcular la temperatura media de cada año



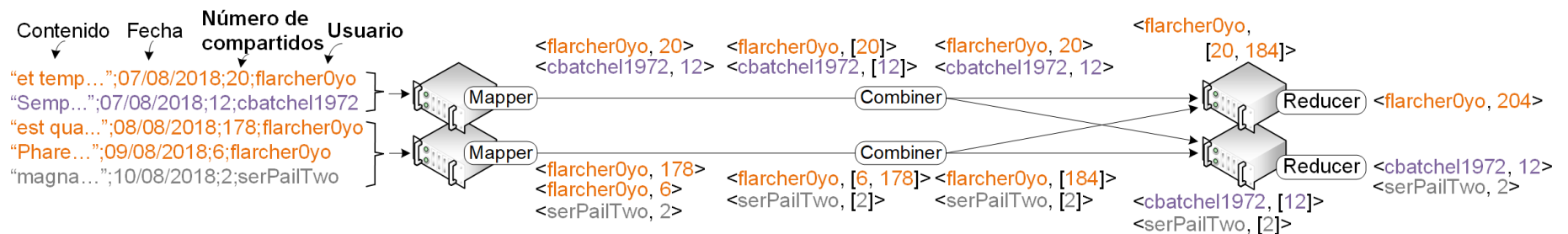
Modelo de procesamiento MapReduce

- 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



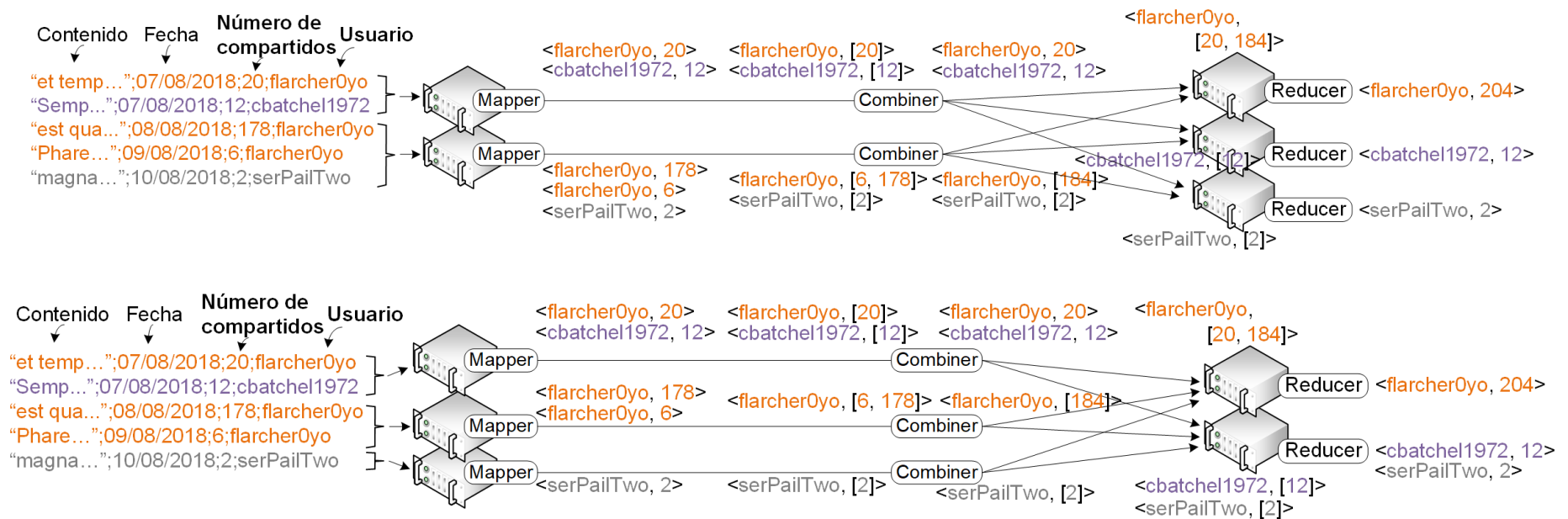
Modelo de procesamiento MapReduce

- 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

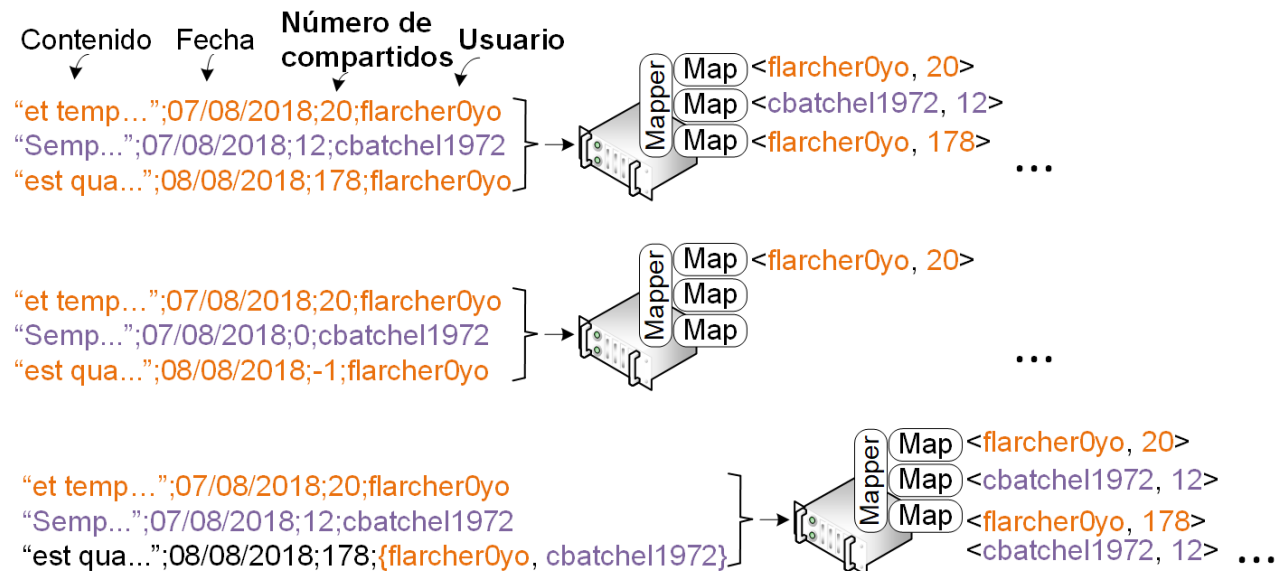


Modelo de procesamiento MapReduce

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



- 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>

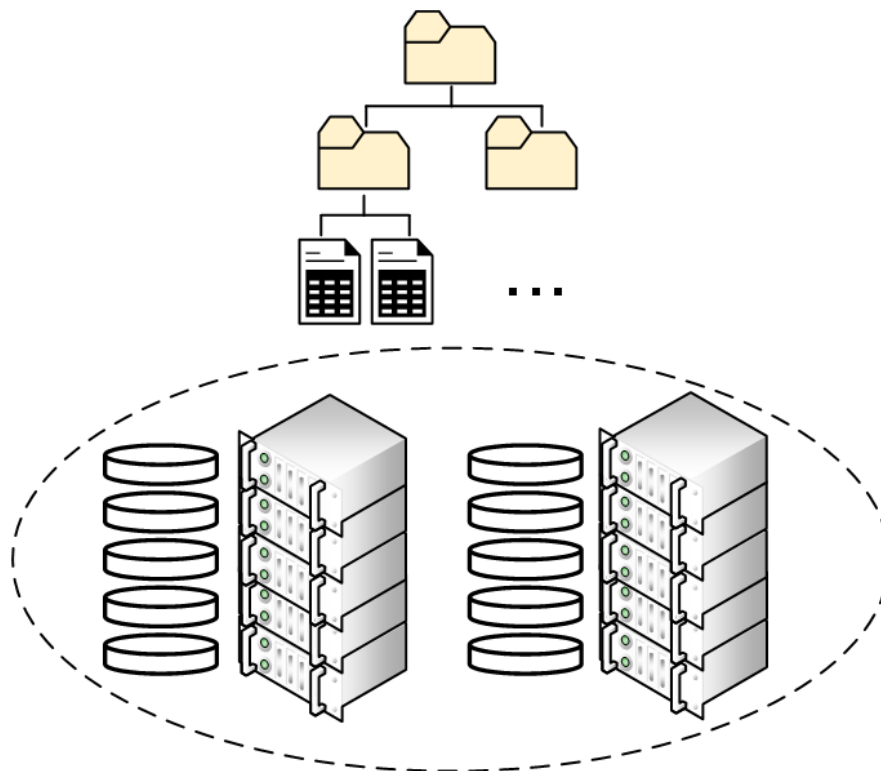
- Reducer: Ejecuta Reduce por cada “subproblema”
 - Reduce puede emitir {0, 1, n} pares <clave, valor>
 - Seleccionar los usuarios que tienen más de 1000 compartidos
 - Seleccionar por cada usuario el número de tweets publicados
 - Seleccionar por cada autor los 10 tweets con más compartidos

- Sencillo
- Distribuido
- Escalable
- Big Data

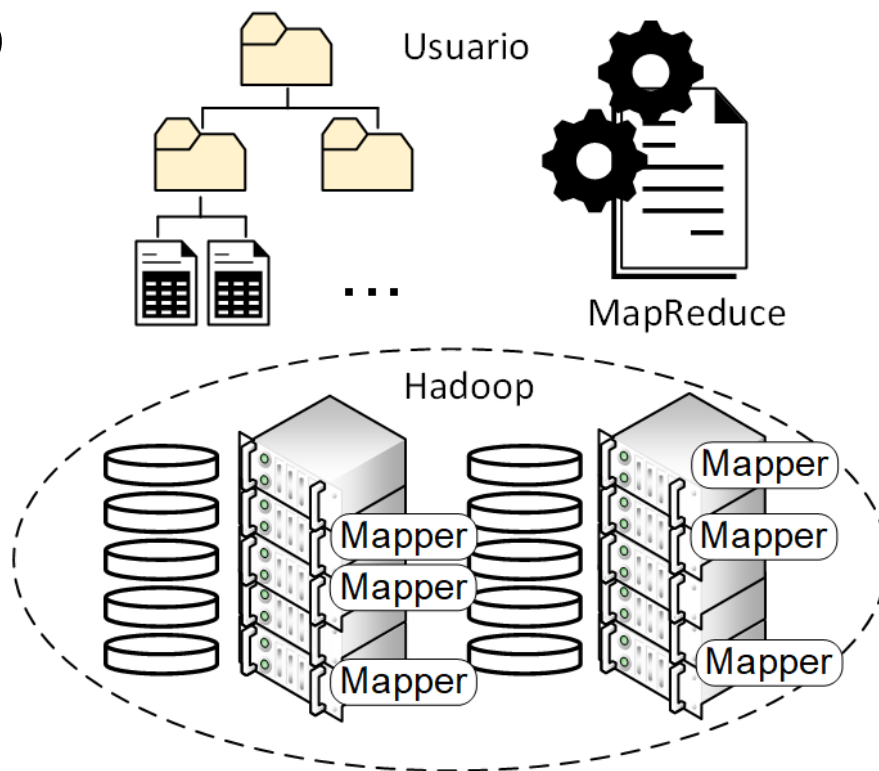
- Frameworks/implementaciones MapReduce:
 - ☐ Google MapReduce
 - ☐ Hadoop MapReduce
 - ☐ Spark
 - ☐ Flink
 - ☐ MongoDB
 - ☐ Riak
 - ☐ ...

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

■ HDFS



■ Hadoop



■ Hadoop

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