Fork-join

Técnicas de Programación Concurrente

Fork-join

Paper de **1963**

"... in an internal **merge sort** all strings in any pass may be created at the same time."

En honor al paper original, implementemos un merge sort concurrente
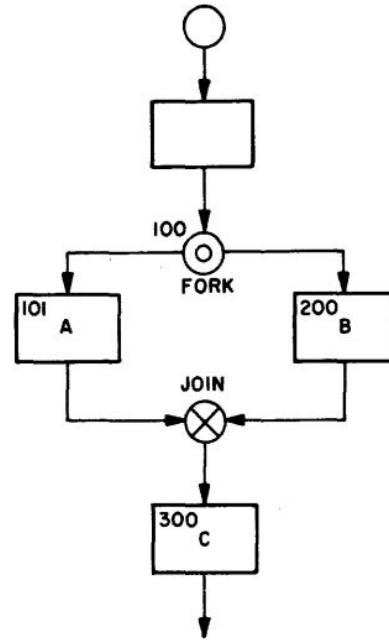

Figure 1. Conventions for drawing fork and join.

Melvin Conway

MapReduce

Paper del año 2004

MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key.

Muchas implementaciones, Hadoop siendo la primera open source

MapReduce
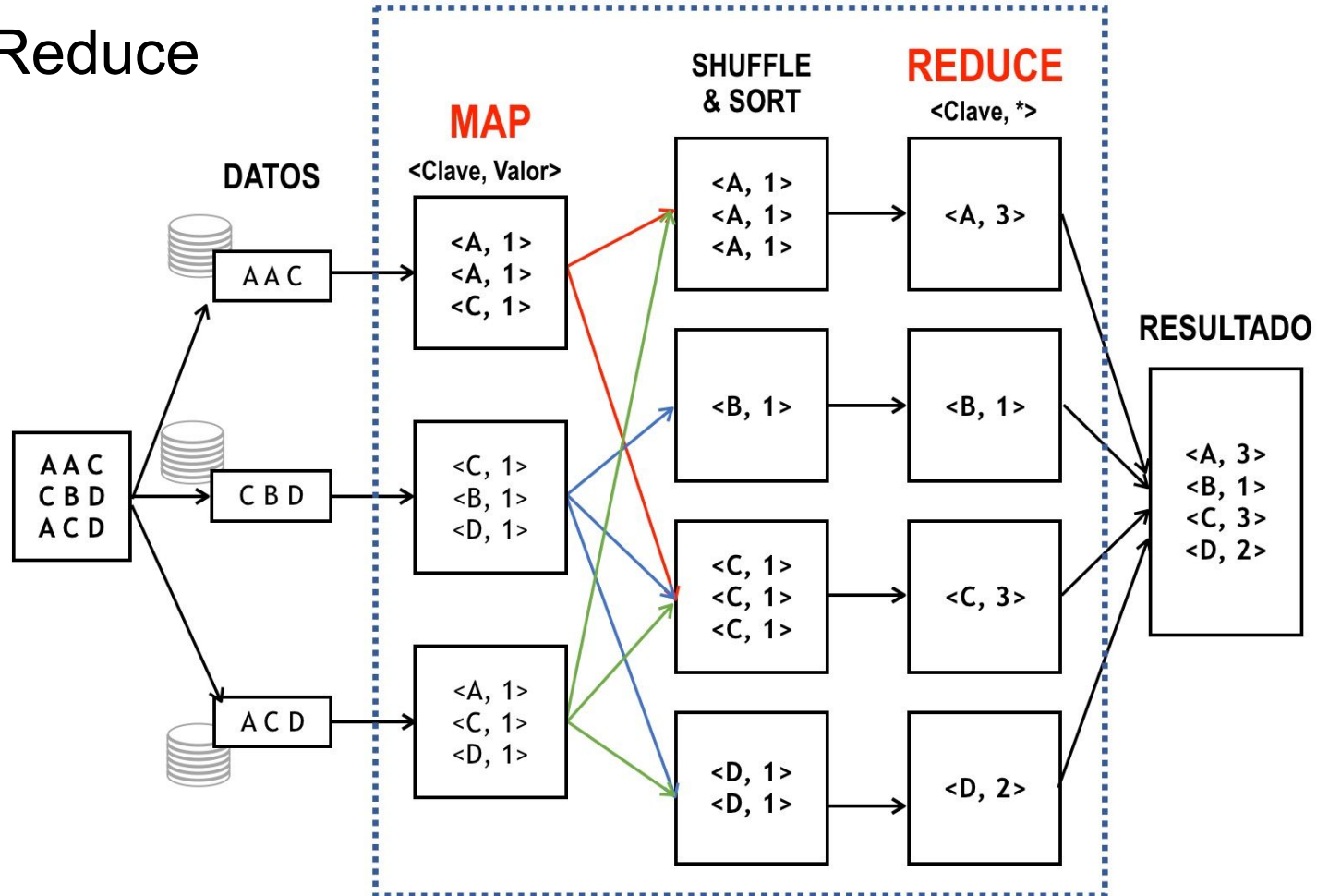
Word count

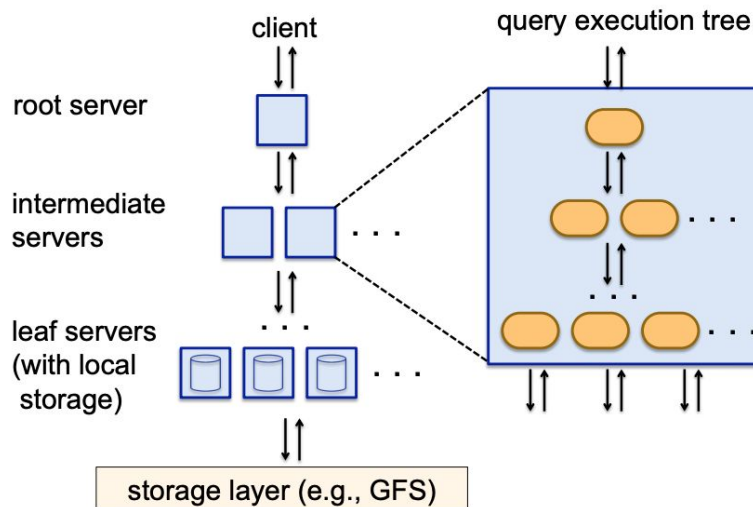Word count es el "Hello World" en el mundo del map reduce

- Map
  - Tomar de potencialmente varios archivos de texto, las lineas
  - De las líneas, las palabras
  - De las palabras sus frecuencias locales
- Reduce
  - Combinar las frecuencias locales en una única tabla con todas las frecuencias

Implementemos el word count!

Dremel

Sistema de consultas en tiempo casi real, expuesto al público como BigQuery

"A query gets pushed down the tree and is rewritten at each step. The result of the query is assembled by aggregating the replies received from lower levels of the tree"

Referencias

- A Multiprocessor System Design, M.E. Conway, 1963 - https://ia600703.us.archive.org/30/items/AMultiprocessorSystemDesignConway1963/A%20Multiprocessor%20System%20Design%20%28Conway%2C%201963%29.pdf
- MapReduce: Simplified Data Processing on Large Clusters, Jeffrey Dean and Sanjay Ghemawat, 2014 - https://static.googleusercontent.com/media/research.google.com/es//archive/mapreduce-osdi04.pdf
- Dremel: Interactive Analysis of Web-Scale Datasets S. Melnik et al., 2010 - https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36632.pdf