



UAB

Universitat Autònoma
de Barcelona

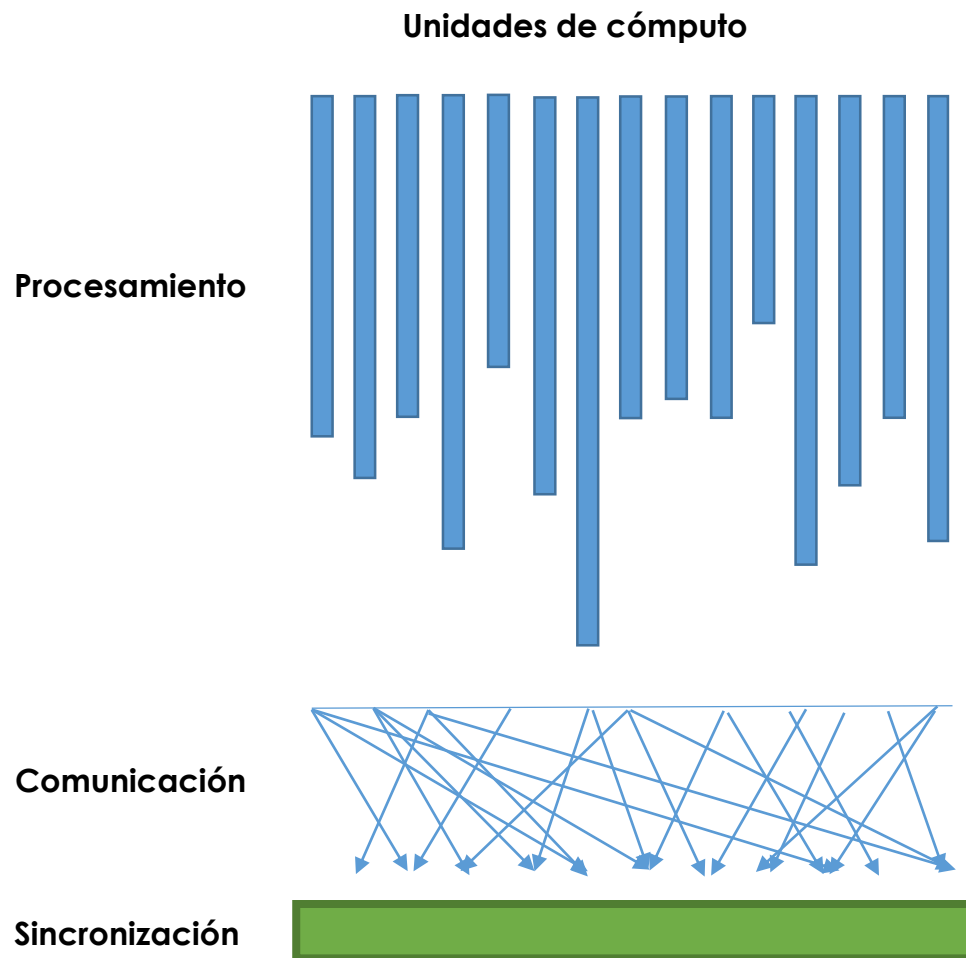
Andrés Cencerrado

Graph processing (II)

Etapas y componentes

Bulk Synchronous Parallel

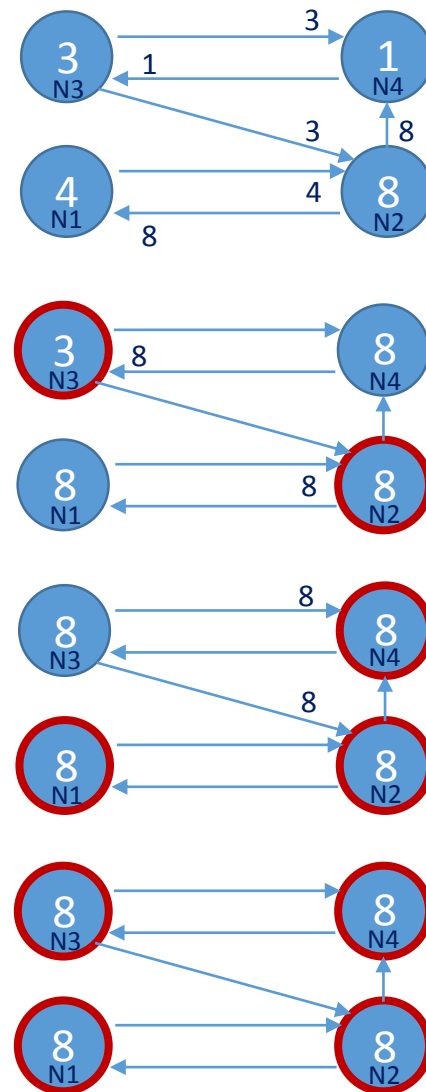
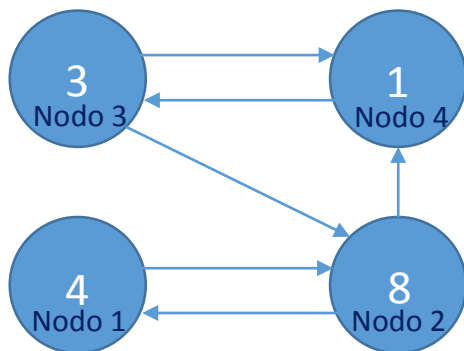
- Paradigma BSP
 1. Procesamiento
 2. Comunicación
 3. Sincronización
- Nodos \equiv unidades de procesamiento
- Este procedimiento se reitera hasta que todos los nodos se quedan sin nada por procesar, por lo que los nodos generalmente mantienen un estado interno que puede ser **activo** o **inactivo (halted)**



Etapas y componentes

Máximo valor

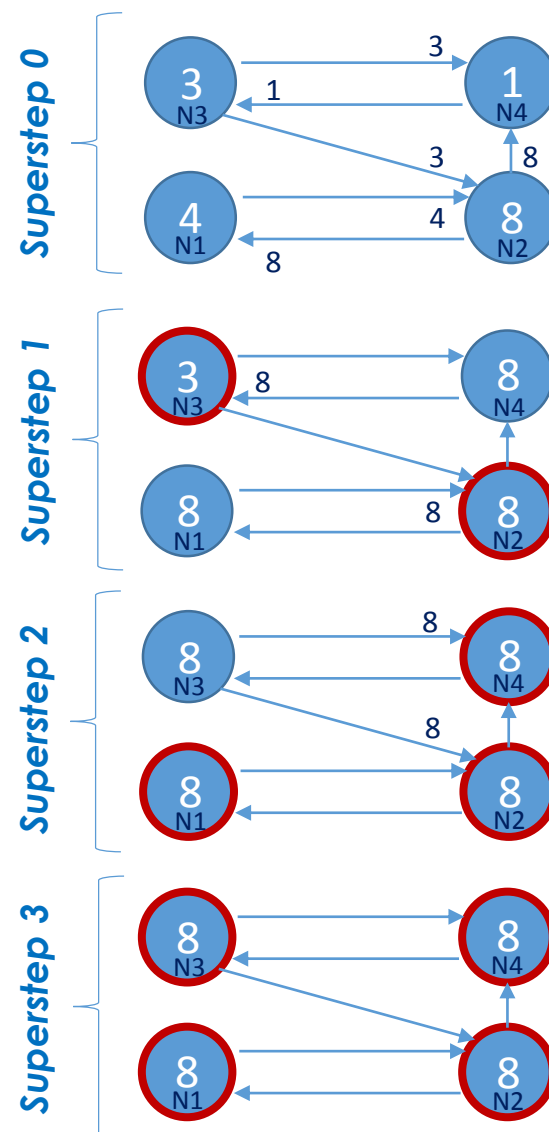
- Paradigma BSP
 1. Procesamiento
 2. Comunicación
 3. Sincronización
- Nodos \equiv unidades de procesamiento
- Este procedimiento se reitera hasta que todos los nodos se quedan sin nada por procesar, por lo que los nodos generalmente mantienen un estado interno que puede ser **activo** o **inactivo (halted)**



Etapas y componentes

Superstep, combiners, aggregators

- **Superstep**
 - Cada iteración que engloba procesamiento, comunicación y sincronización
- **Combiners**
 - Procedimientos “ligeros” previos al procesamiento principal de cada nodo, aplicados a los mensajes entrantes
- **Aggregators**
 - Información global; permiten realizar cambios en el flujo de ejecución



Características

MapReduce no es muy adecuado para grafos...

- **Dos limitaciones principales de MapReduce:**
 1. **No es adecuado para computación iterativa**
 2. **Imposibilidad de transmitir mensajes**
- 1. Algoritmos de grafos normalmente son iterativos, y algunos de ellos requieren una gran cantidad de iteraciones.
MapReduce solo realiza una iteración, para manejar múltiples iteraciones, hay que desarrollar el *framework* aparte.
Generalmente esto implica gran E / S y un tiempo de inicio de trabajo adicional, por lo que es muy poco eficiente.
- 2. Paso de mensajes es básico en *graph processing*.
PageRank, por ejemplo, para calcular el peso de un vértice, requiere conocer las puntuaciones de todos sus vecinos. Este cálculo requiere el mensaje de sus vecinos.



Características

Ventajas a considerar

- **Rendimiento**
 - Manejo intensivo de relaciones → mejora de rendimiento en varios órdenes de magnitud respecto a BBDD relacionales
- **Flexibilidad**
 - Gran capacidad de adaptarse a cambios en el modelo (adición de nuevas entidades, modificación de relaciones, etc.)
- **Versatilidad, no interferencia**
 - Evolución y cambio de modelo sin alterar esquema empresarial (o del *Business Intelligence*)
 - Motores modernos incluyen optimizaciones para ETL, análisis y presentación



Herramientas

- Herramientas de Graph processing



A green-tinted image of a tablet displaying the word "MOOC" in large, bold, white letters.

MOOC

Graph processing (II)

A green-tinted image of a laptop keyboard with the UAB MOOC logo overlaid. The logo includes the UAB logo, the word "MOOC", and the text "Escola de Postgrau".

UAB
Universitat Autònoma de Barcelona

MOOC
Escola de
Postgrau

UAB

Universitat Autònoma
de Barcelona