

## TP3 - Sistemas Distribuidos

Daniel J. Foguelman -> Guido Chari

DC - FCEyN - UBA

Sistemas Operativos, 1c - 2014

## Definición

"Map-reduce is a **programming model** for expressing **distributed computations** on **massive amounts of data** and an execution framework for **large-scale data processing** on clusters of **commodity servers**"

<http://research.google.com/archive/mapreduce.html>

## ¿Qué es?

- Un modelo de desarrollo y su implementación.
- Diseñado para el cómputo de datos de gran escala.
- Donde la performance es un un aspecto importante.
- Con una arquitectura de bajo costo (depende la implementación).
- Y un requerimiento fuerte: Escalabilidad lineal (doble procesamiento, mitad del tiempo).

Conceptos importantes abstraídos para el programador:

- Paralelización y distribución de daots.
- Abstracción de la capa de comunicación.
- Tolerancia a fallos.
- I/O scheduling.
- Balance de carga.

¡Datos, Datos, Datos !

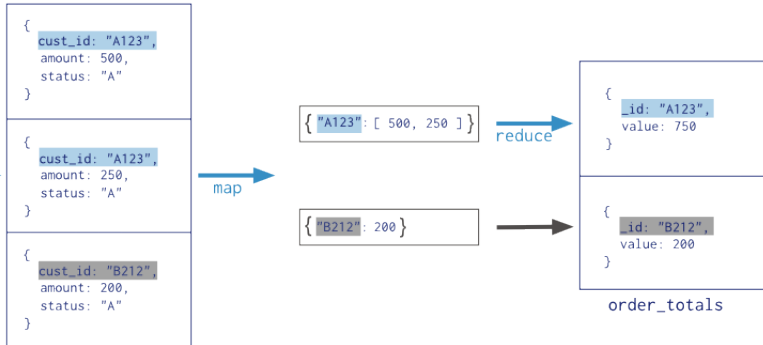
Google procesa 24  
petabytes/día

## Map - Reduce

```
> map (in_key, in_value) ->  
    list(out_key, intermediate_value)  
  
> reduce (out_key, list(intermediate_value)) ->  
    list(out_value)
```

Orientado a documentos, clave-valor.

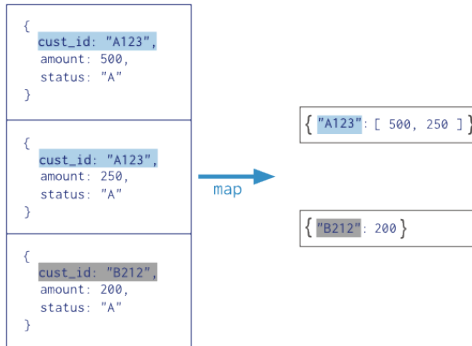
## Un ejemplo



<http://docs.mongodb.org/manual/core/map-reduce>

# Map

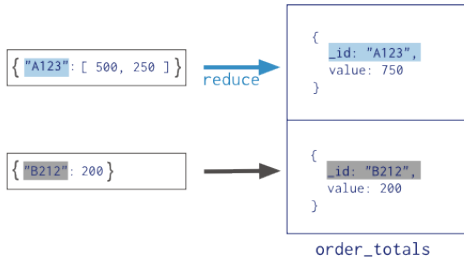
```
map = function() {  
    emit(this.cust_id, this.amount);  
}
```



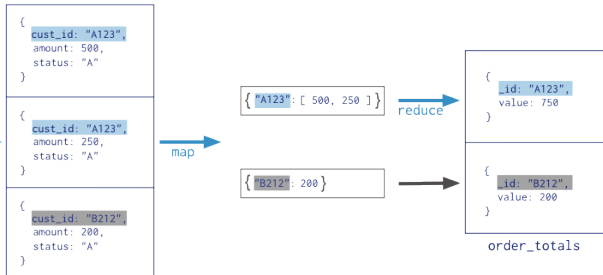


## Reduce

```
reduce = function(key, values) {  
    return Array.sum(values);  
}
```

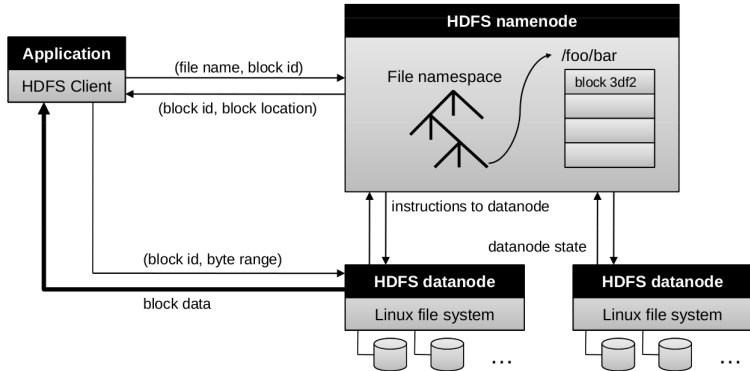


```
> use sales
> show collections
  orders
> db.orders.mapReduce(map, reduce, order_totals);
> show collections
  orders
  order_totals
```



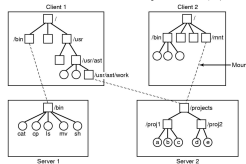
Localización de los datos para minimizar I/O sobre la red, brindando:

- Replicación
- Disponibilidad



- El *Network File System* es un protocolo que permite acceder a FS remotos como si fueran locales, utilizando RPC.
- La idea es que un FS remoto se monta en algún punto del sistema local y las aplicaciones acceden a archivos de ahí, sin saber que son remotos.
- Para poder soportar esto, los SO incorporan una capa llamada *Virtual File System*.
- Esta capa tiene *vnodes* por cada archivo abiertos. Se corresponden con inodos, si el archivo es local. Si es remoto, se almacena otra información.
- Así, los pedidos de E/S que llegan al VFS son despachados al FS real, o al *cliente de NFS*, que maneja el protocolo de red necesario.
- Si bien del lado del cliente es necesario un módulo de kernel, del lado del server alcanza con un programa común y corriente.

### Network File System (1)



- Examples of remote mounted file systems
- Directories are shown as squares, files as circles

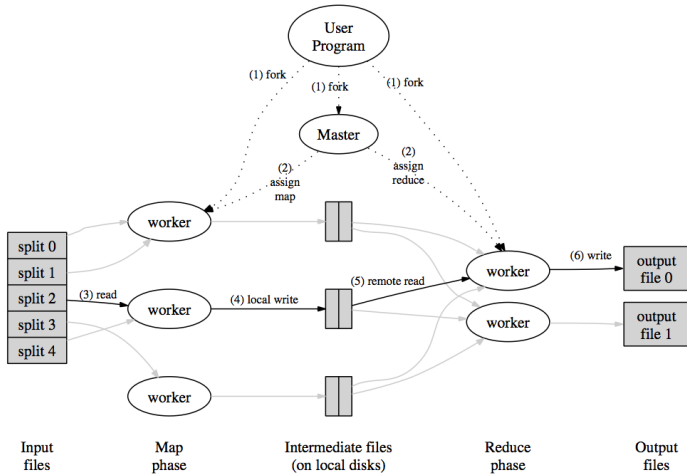
- Otros FS distribuidos funcionan de manera similar (en cuanto a su integración con el kernel).
- Notar que, desde cierto punto de vista, NFS no es 100 % distribuido, ya que todos los datos de un mismo “directorio” deben vivir físicamente en el mismo lugar.
- Hay FS 100 % distribuidos, como AFS o Coda.
- Han tenido un éxito relativo.

## Framework de ejecución

- Map
- Combine
- Reduce

¡Alguien debe orquestar la ejecución!

[http://www.cs.berkeley.edu/~matei/talks/2009/msr\\_mapreduce\\_scheduling.pdf](http://www.cs.berkeley.edu/~matei/talks/2009/msr_mapreduce_scheduling.pdf)



Una base de datos No-SQL que implementa:

- MapReduce.
- Replicación y alta disponibilidad.
- Autosharding, escalabilidad horizontal.
- GridFS.



Se espera que:

- Aprendan a manejar el modelo de programación.
  - ¡Como algo uan pero un poquito más!
- Realicen un análisis de arquitectura.
- Entiendan los criterios de análisis para observar un sistema distribuido.
- Todo en una VM.

- Data-Intensive Text Processing with MapReduce, Jimmy Lin and Chris Dyer University of Maryland, College Park
- Mining of Massive Datasets Anand Rajaraman Jure Leskovec Stanford Univ. Jeffrey D. Ullman
- Data-Intensive Text Processing with MapReduce Jimmy Lin and Chris Dyer University of Maryland, College Park
- The Little MongoDB Book Karl Seguin
- <http://docs.mongodb.org/manual/>
- <http://research.google.com/archive/mapreduce-osdi04-slides/index.html>
- <http://static.googleusercontent.com/media/research.google.com/es//archive/mapreduce-osdi04.pdf>
- <http://www.mongodb.com/dl/big-data>
- [http://info.mongodb.com/rs/mongodb/images/MongoDB-Performance-Considerations\\_2.4.pdf](http://info.mongodb.com/rs/mongodb/images/MongoDB-Performance-Considerations_2.4.pdf)
- <http://christophermeiklejohn.com/distributed/systems/2013/07/12/readings-in-distributed-systems.html>

