Diseño de Sistemas Distribuidos

Máster en Ciencia y Tecnología Informática

Curso 2018-2019

Sistemas escalables en entornos distribuidos

Alejandro Calderón Mateos, acaldero@inf.uc3m.es

Óscar Pérez Alonso, oscar@lab.inf.uc3m.es

Félix García Carballeira fgcarball@inf.uc3m.es

Contenidos

Big
Big
Big
Big
Big

- Inicios

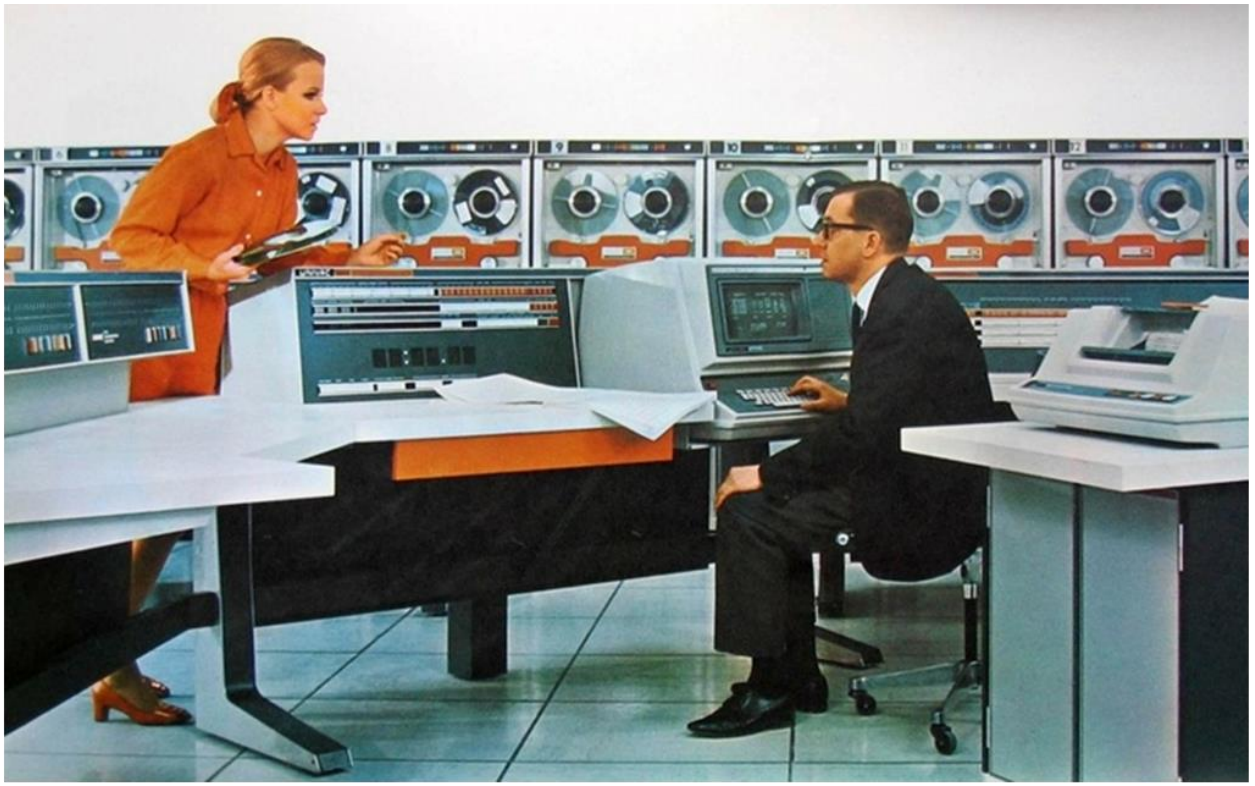
http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838

Los inicios: sistemas **no** distribuidos…

Centralizado

https://www.dcsorg.com/images/image_centralized_management.jpg

Los inicios: sistemas centralizados…



~ 1960 →

http://www.nethistory.info/History%20of%20the%20Internet/origins.html#apps

Los inicios: sistemas centralizados…

- Sistema centralizado
  - Mejor mantenimiento
  - Compartición de recursos
  - Repartición de costes

~ 1960

http://www.nethistory.info/History%20of%20the%20Internet/origins.html#apps

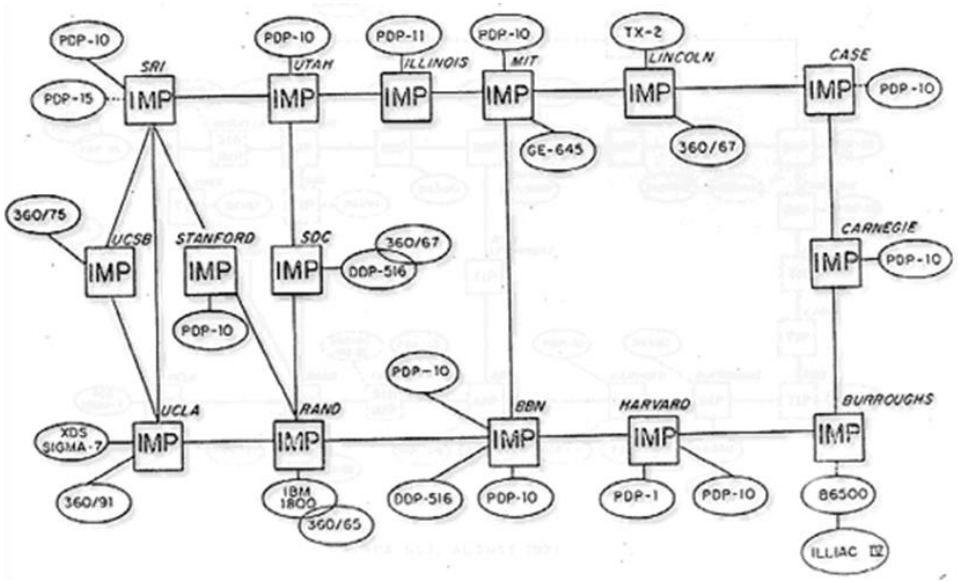
Los inicios: ARPANET…

ARPANET

~ 1969

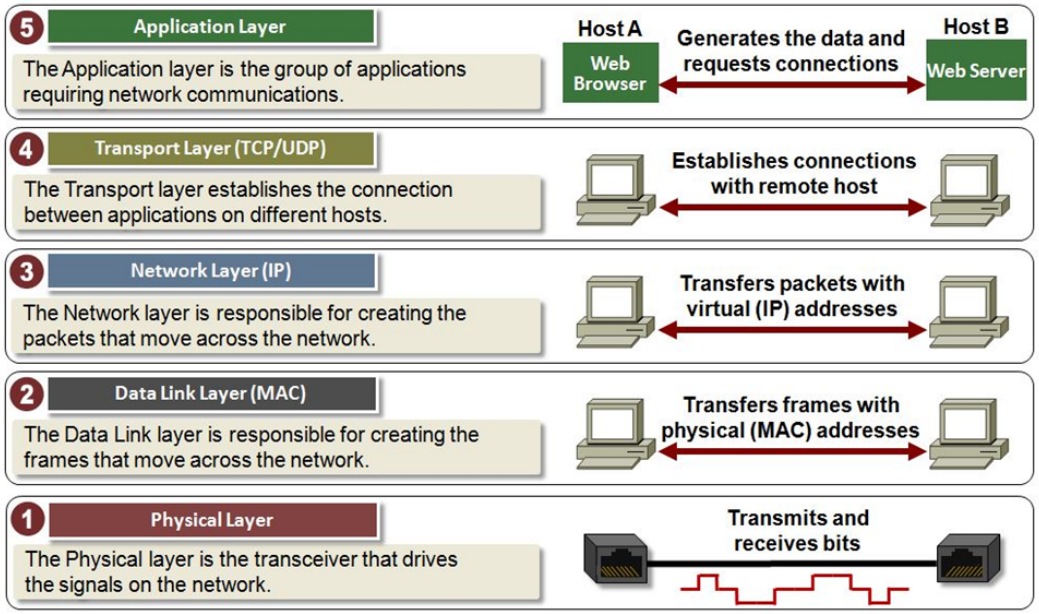
http://www.nethistory.info/History%20of%20the%20Internet/origins.html#apps

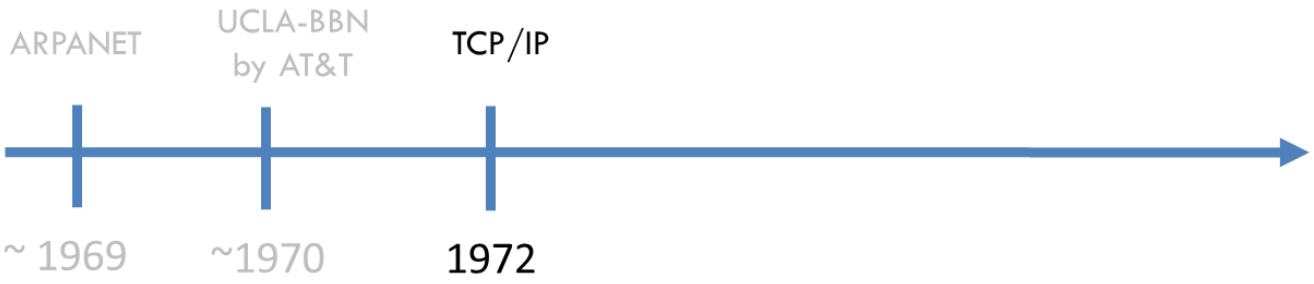
Los inicios: pre-Internet…



http://www.nethistory.info/History%20of%20the%20Internet/origins.html#apps

Los inicios: TCP/IP…

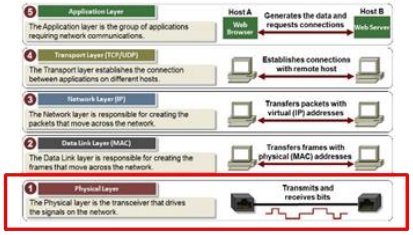


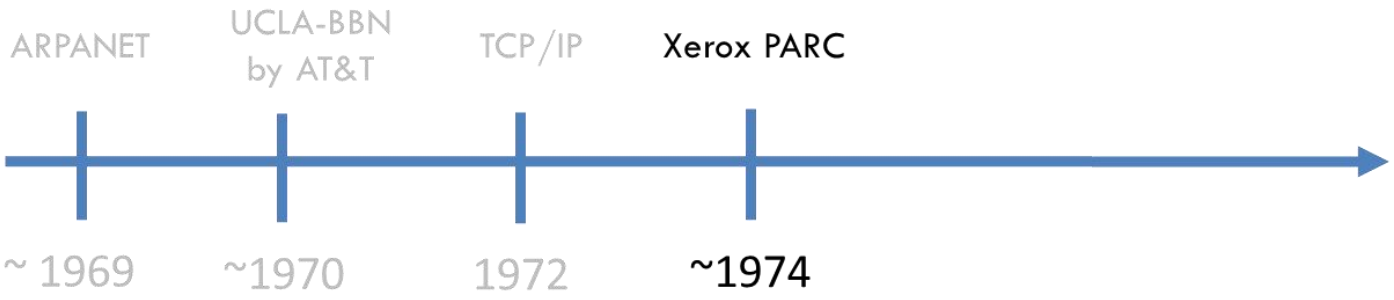


http://microchip.wdfiles.com/local--files/tcpip:tcp-ip-five-layer-model/TCPIP_5_layer_overview.JPG

Los inicios: Ethernet…

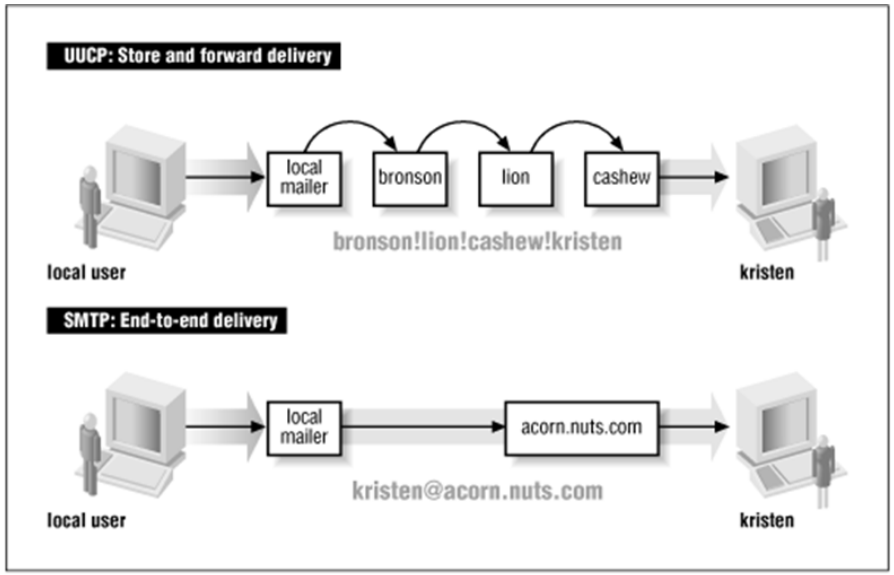
Pin	Signal ID	T568A Color	T568B Color	Pins on plug face (socket is reversed)
1	TX+	white/green stripe	white/orange stripe	
2	TX-	green solid	orange solid	
3	RX+	white/orange stripe	white/green stripe	
4	7 - 30VDC	blue solid	blue solid	
5	7 - 30VDC	white/blue stripe	white/blue stripe	
6	RX-	orange solid	green solid	
7	GROUND	white/brown stripe	white/brown stripe	
8	GROUND	brown solid	brown solid	

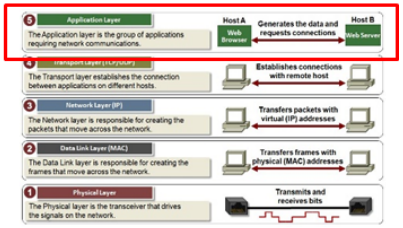


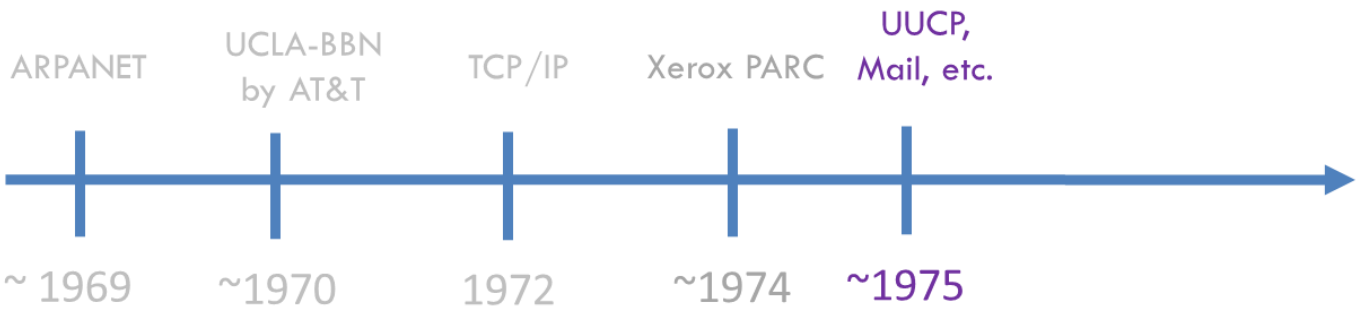


http://www.nethistory.info/History%20of%20the%20Internet/origins.html#apps

Los inicios: primeros servicios en red…

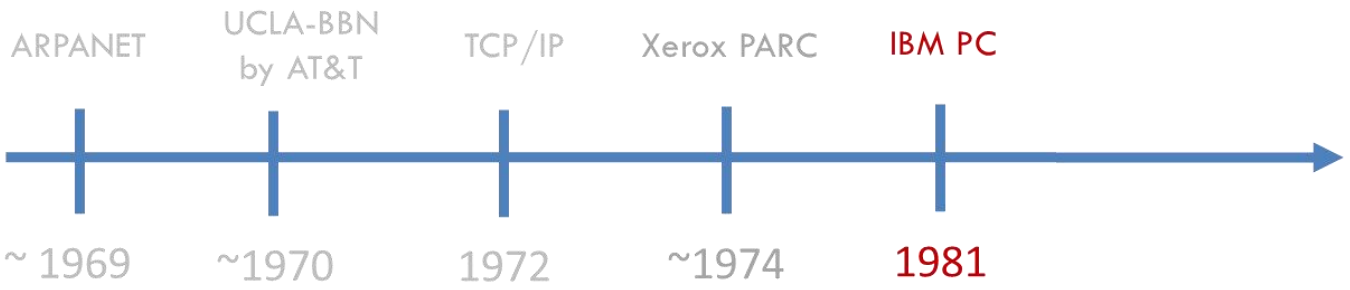






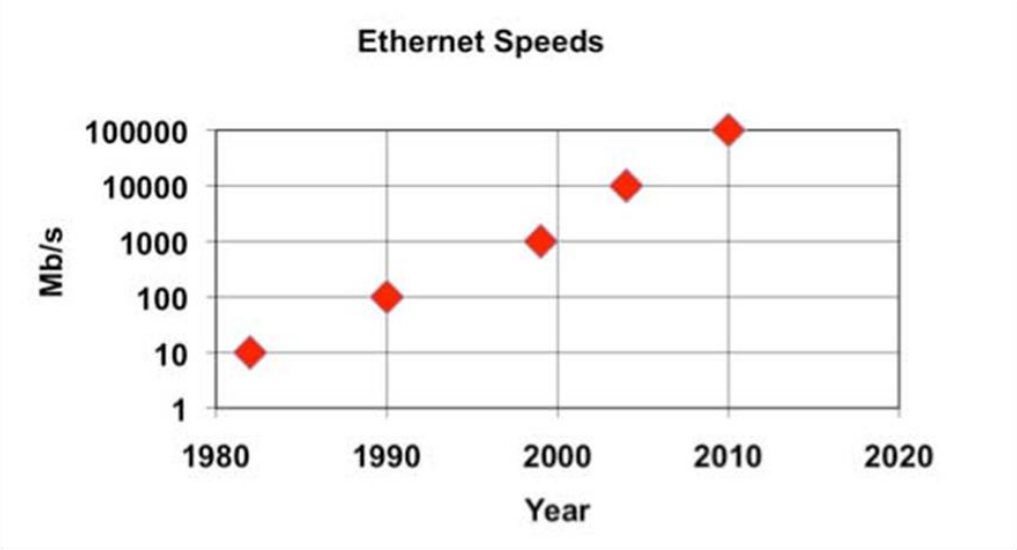
http://edc.tversu.ru/elib/inf/0091/tcpip/figs/tcp2_0303.gif

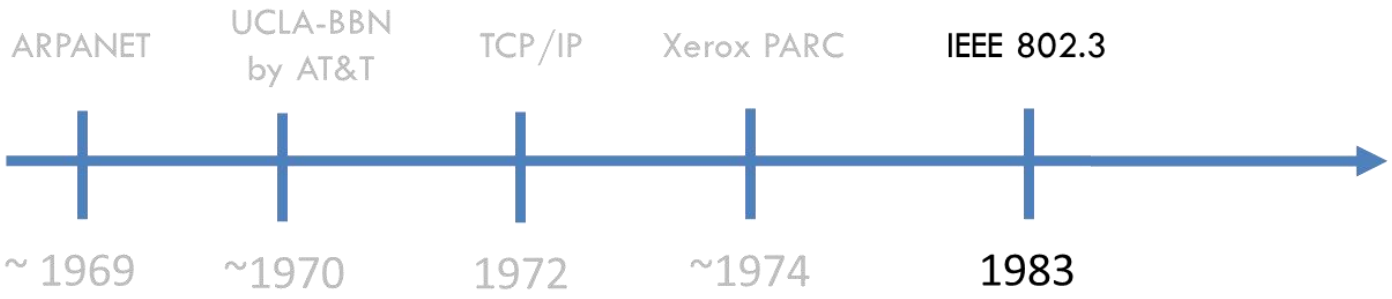
Los inicios: un ordenador en cada…



https://en.wikipedia.org/wiki/IBM_Personal_Computer

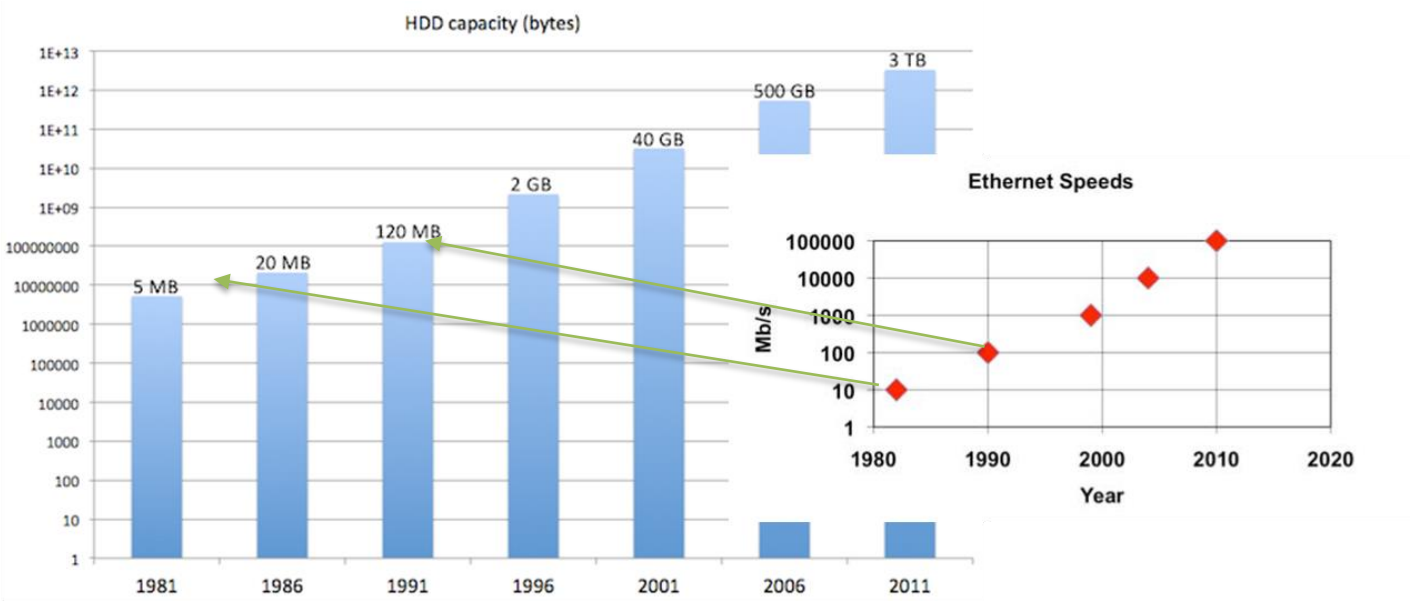
Los inicios: Ethernet más rápida…

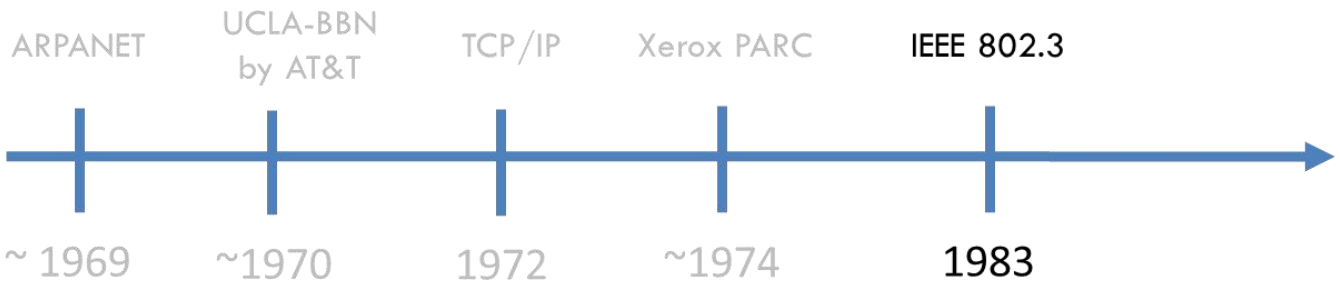




http://www.thefoa.org/tech/ref/appln/OLAN.html

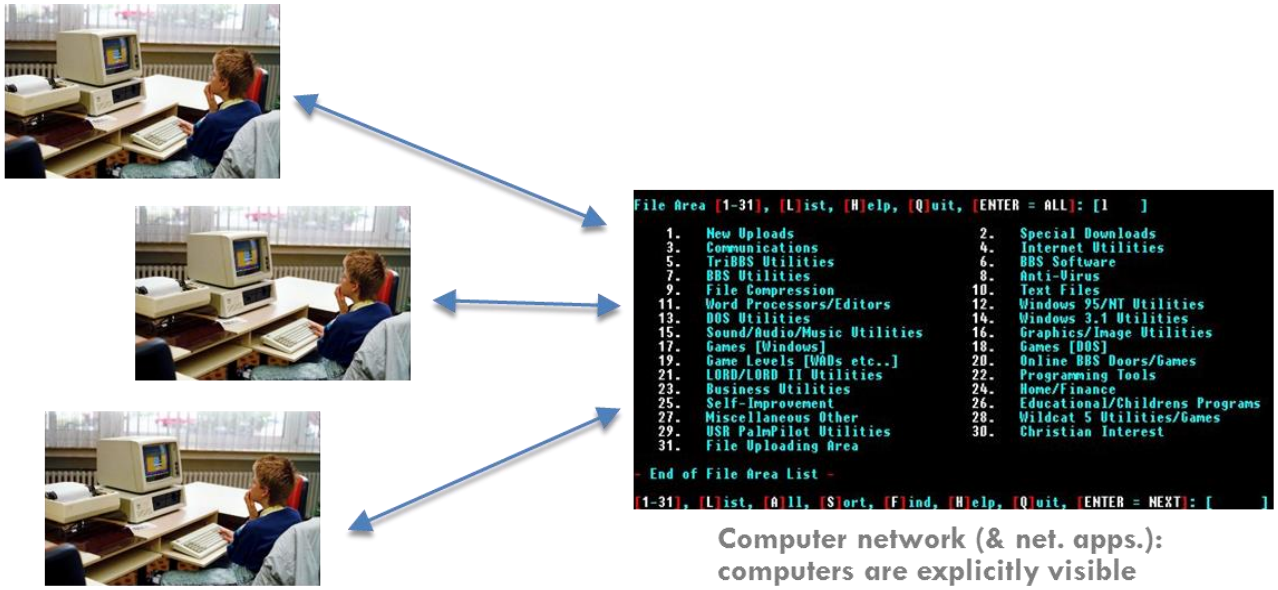
Los inicios: un disco duro en segundos…



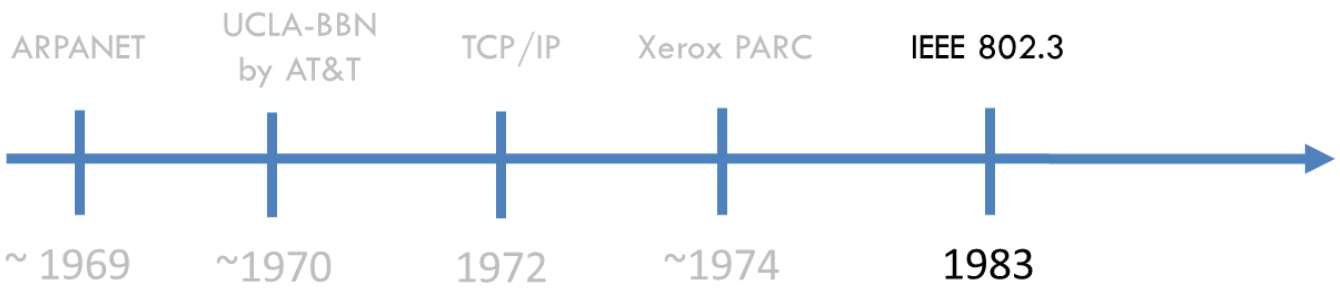


http://cdn.arstechnica.net/2011/09/23/hdd-capacity-scale-4e7ce6c-intro.png

Los inicios: compartir información…



Computer network (& net. apps.): computers are explicitly visible



https://bitcointalk.org/index.php?topic=430357.0

Contenidos

Big
Big
Big
Big
Big

- Inicios
- Transición

http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838

Transición…

Virtually centralized ☺

Centralized

Distributed

Distributed system:
existence of multiple elements is transparent

https://www.dcsorg.com/images/image_centralized_management.jpg

Las transiciones no son fáciles…

Virtually centralized ☺

Centralized

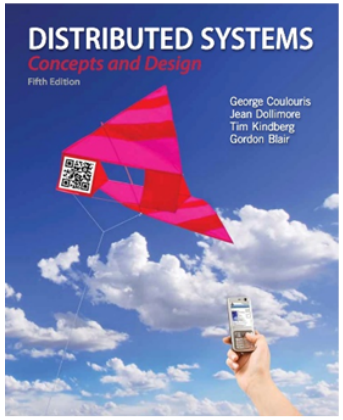
Distributed

Distributed system:
existence of multiple elements is transparent

https://www.dcsorg.com/images/image_centralized_management.jpg

Las transiciones no son fáciles…

Expectativas

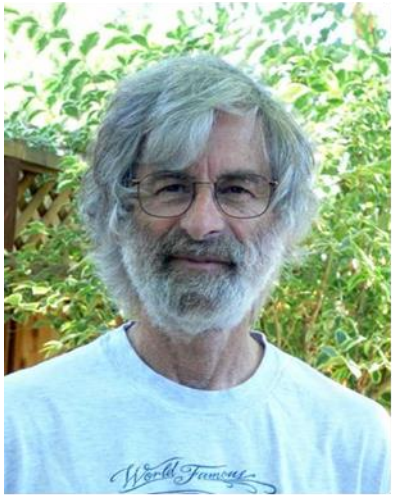


"We define a distributed system as one in which hardware or software components located at networked computers communicate and coordinate their actions only by passing messages"

"A distributed system is a collection of independent computers that appears to its users as a single coherent system"

Las transiciones no son fáciles…

Realidades



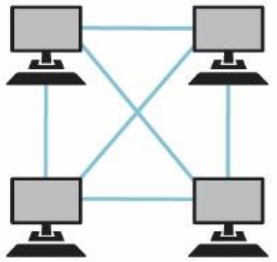
- “A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable”

Leslie Lamport

Las transiciones no son fáciles…

Retos…

Peter Deutsch et al.

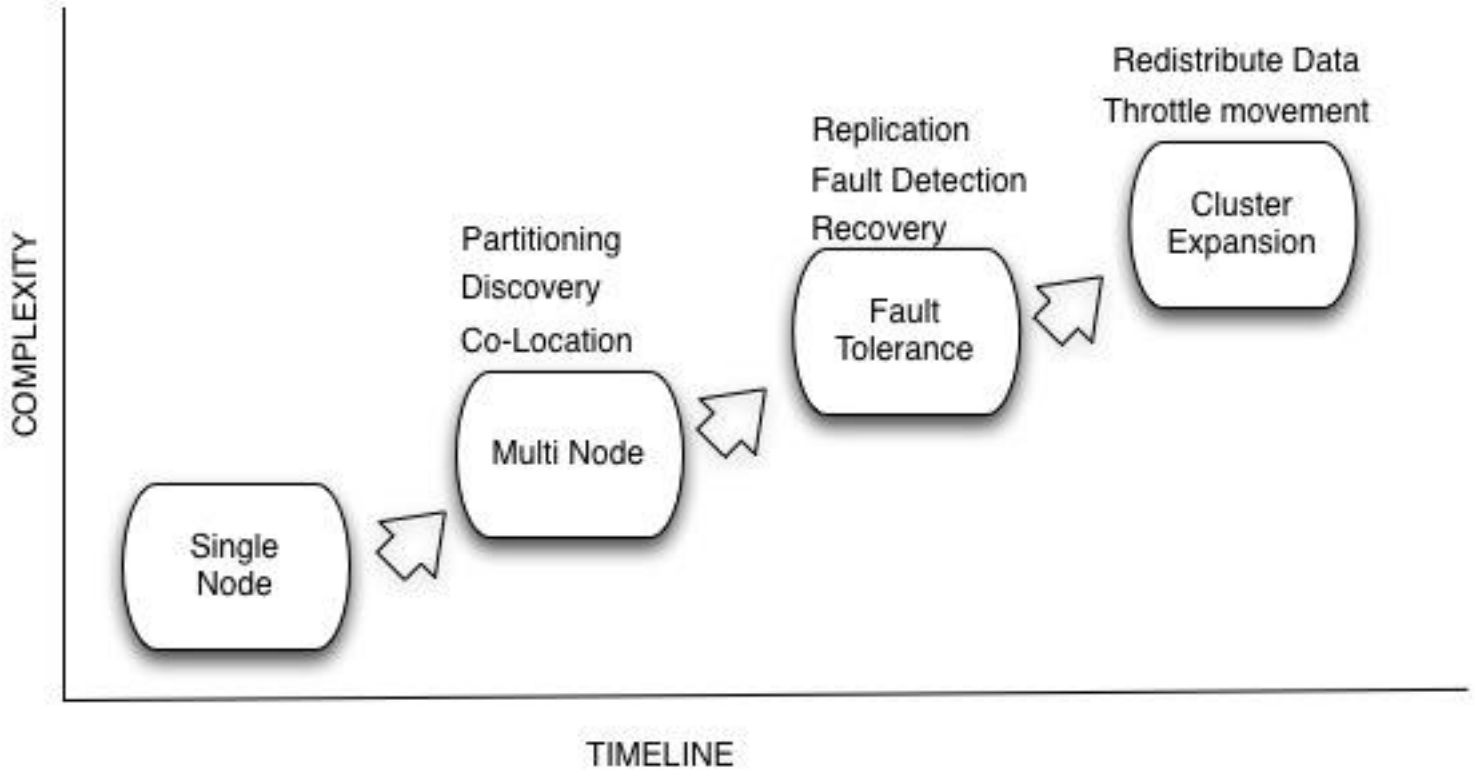


- The network is reliable
- The network is secure
- The network is homogeneous
- The topology does not change
- Latency is zero
- Bandwidth is infinite
- Transport cost is zero
- There is one administrator

http://cdn.comsol.com/wordpress/2014/02/Speeding-up-communications-distributed-memory-computing-copy.jpg

Las transiciones no son fáciles…

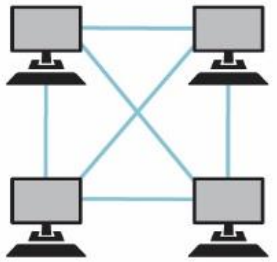
Retos…



http://thenewstack.io/helix-a-linkedin-framework-for-distributed-systems-development/

Las transiciones no son fáciles…

Áreas de trabajo…



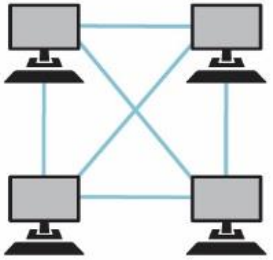
- Heterogeneity
- Openness
- Security
- Scalability
- Failure Handling
- Concurrency
- Transparency

http://cdn.comsol.com/wordpress/2014/02/Speeding-up-communications-distributed-memory-computing-copy.jpg

Las transiciones no son fáciles…

Lecciones aprendidas…

- Heterogeneity
- Openness
- Security
- Scalability
- Failure Handling
- Concurrency
- Transparency

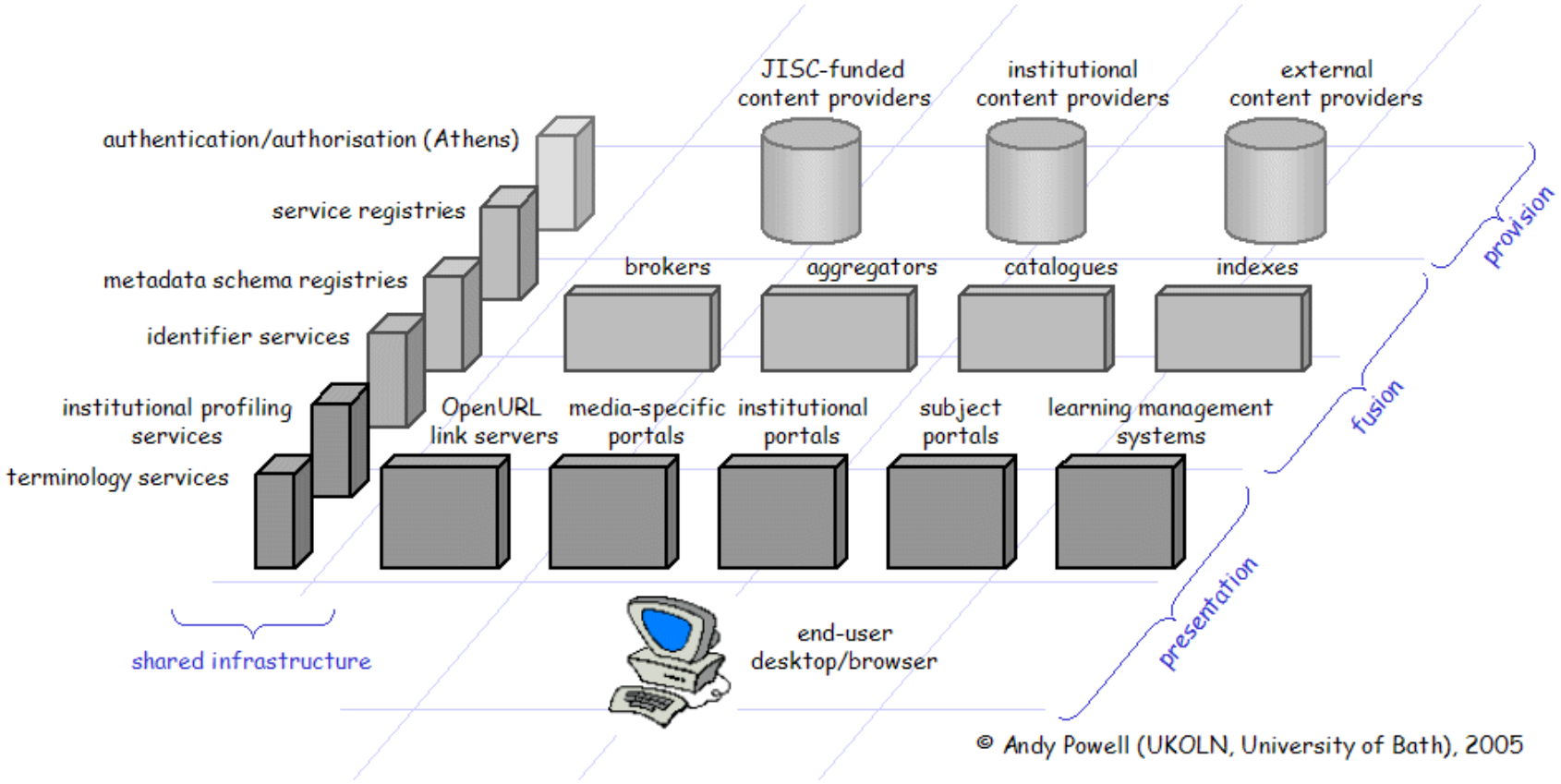


Software/hardware extra en S.D.: conseguir que la existencia de múltiples elementos sea transparente

http://www.pixempire.com/images/preview/orchestra-director-with-stick-icon.jpg

Las transiciones no son fáciles…

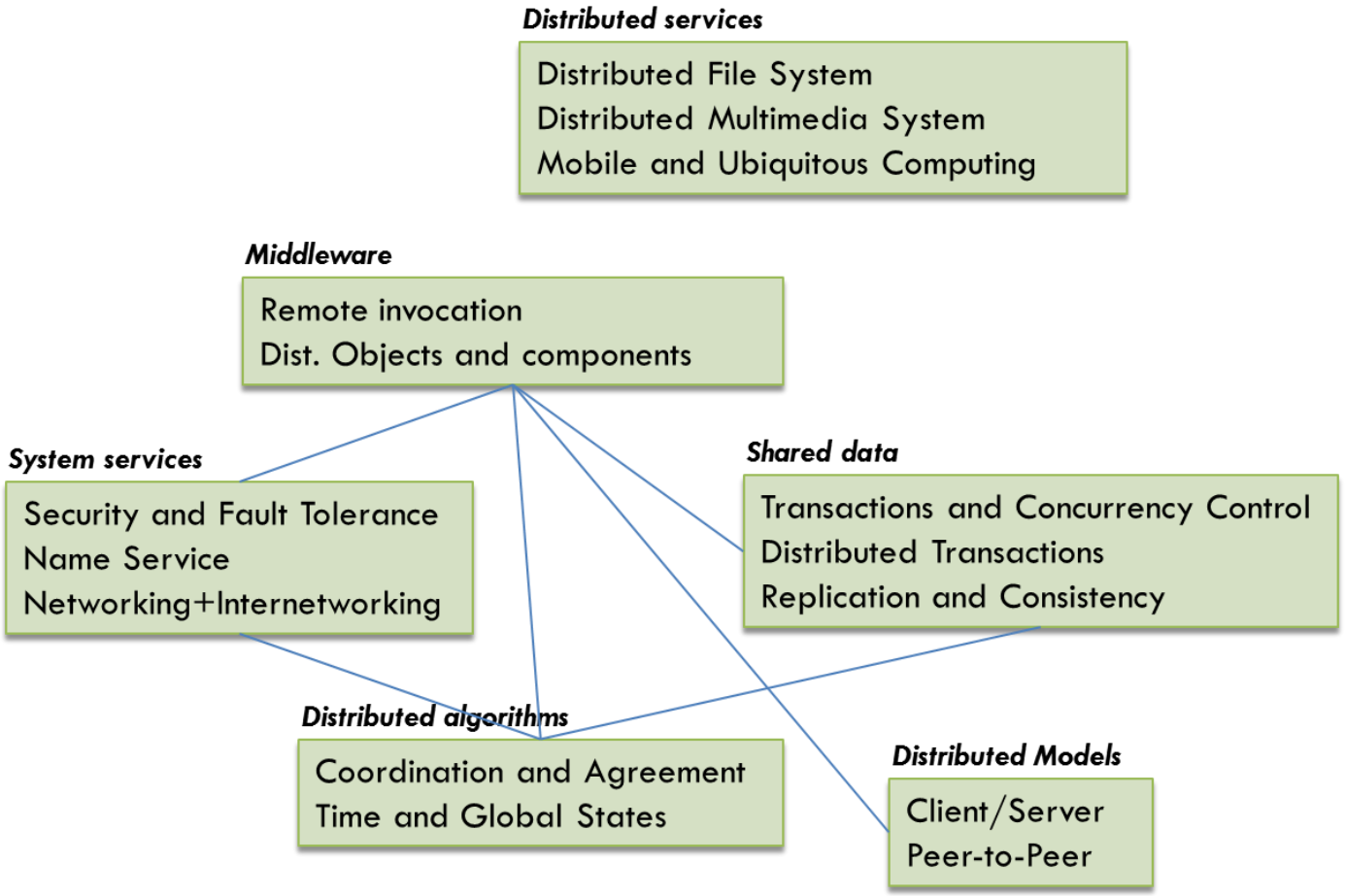
Lecciones aprendidas…



http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/

Las transiciones no son fáciles…

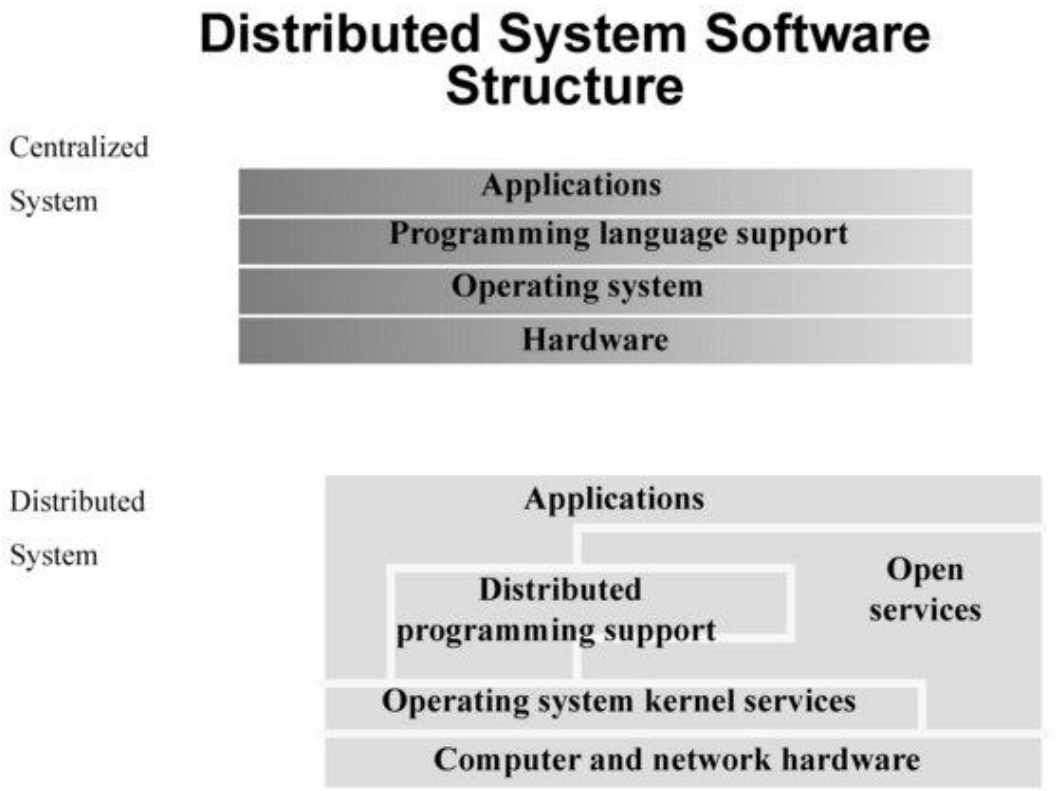
Lecciones aprendidas: elementos típicos…



http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/

Las transiciones no son fáciles…

Lecciones aprendidas: arquitectura típica…



http://books.cs.luc.edu/distributedsystems/issues.html

Contenidos

Big
Big
Big
Big
Big

- Inicios
- Transición
- (re)Evolución

http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838

Sistemas distribuidos: mayor tamaño…

Distributed

Distributed system:
existence of multiple elements is transparent

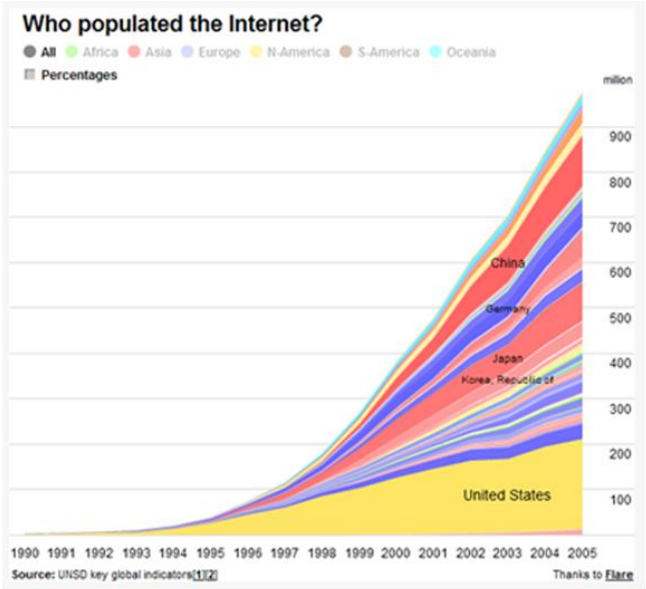
Cloud

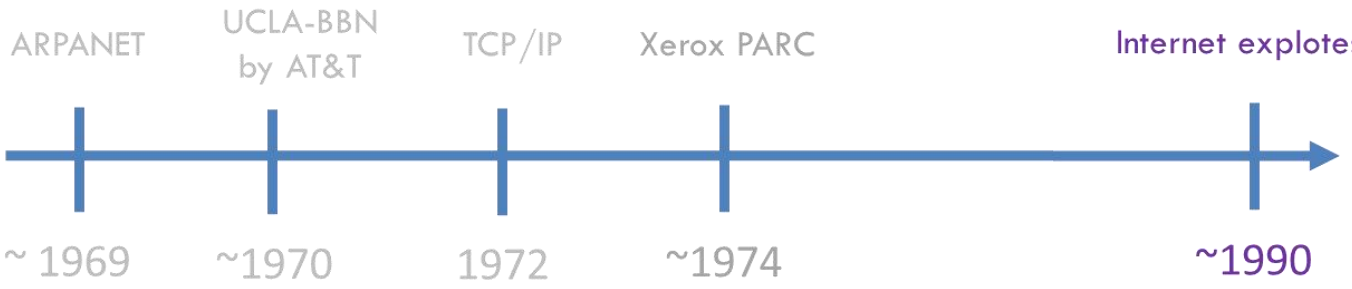
Cloud computing:
Ubiquitous, on-demand access to shared pool of computing resources

https://www.dcsorg.com/images/image_centralized_management.jpg
http://www.hsi.es/images/cc.png

Sistemas distribuidos: mayor tamaño…

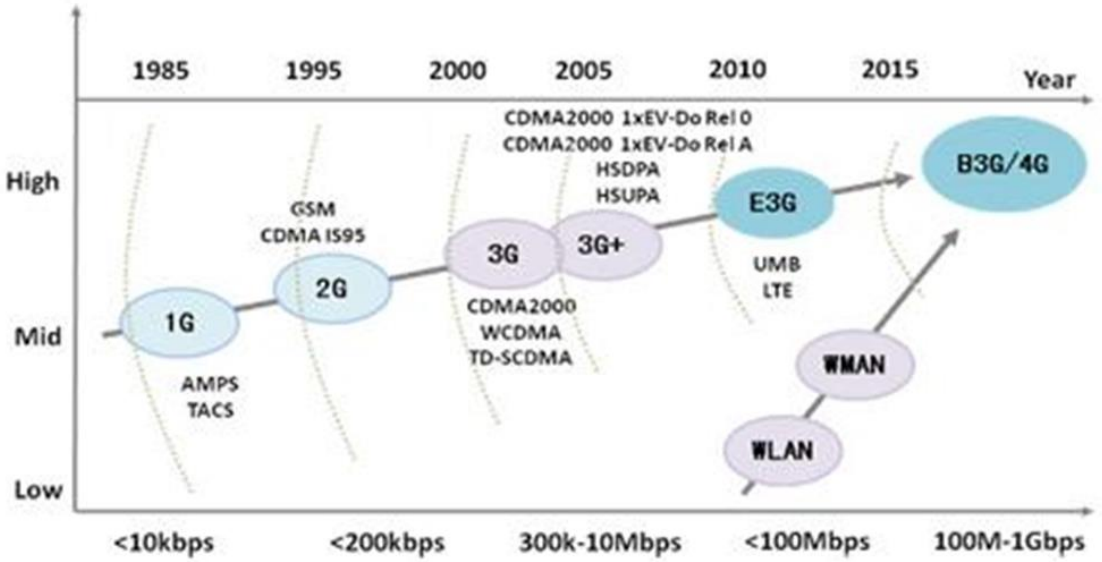
Todas las **casas** conectadas

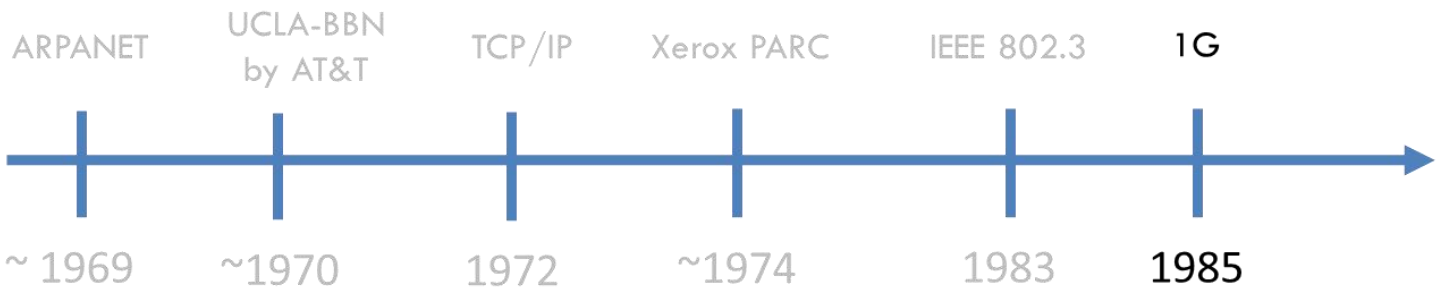




https://bitcointalk.org/index.php?topic=430357.0

Sistemas distribuidos: movilidad…

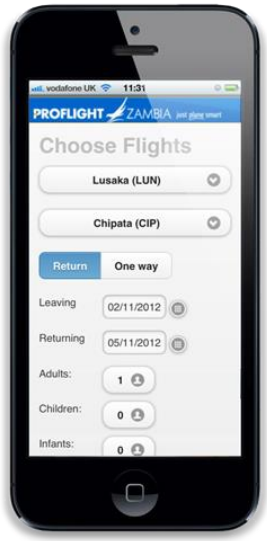


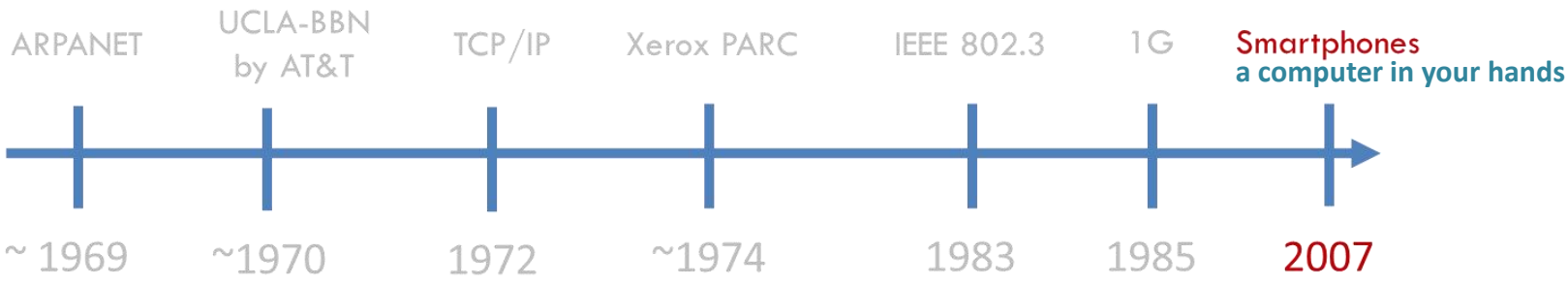


http://img.frbiz.com/news/145317_s/Mobile_communication_base_station_radio_equipment_greet_with_the_explosive_growth_mobile_communication_base_station_radio_equipment_3G_communication_industry.jpg

Sistemas distribuidos: mayor tamaño…

Todas las **personas** conectadas

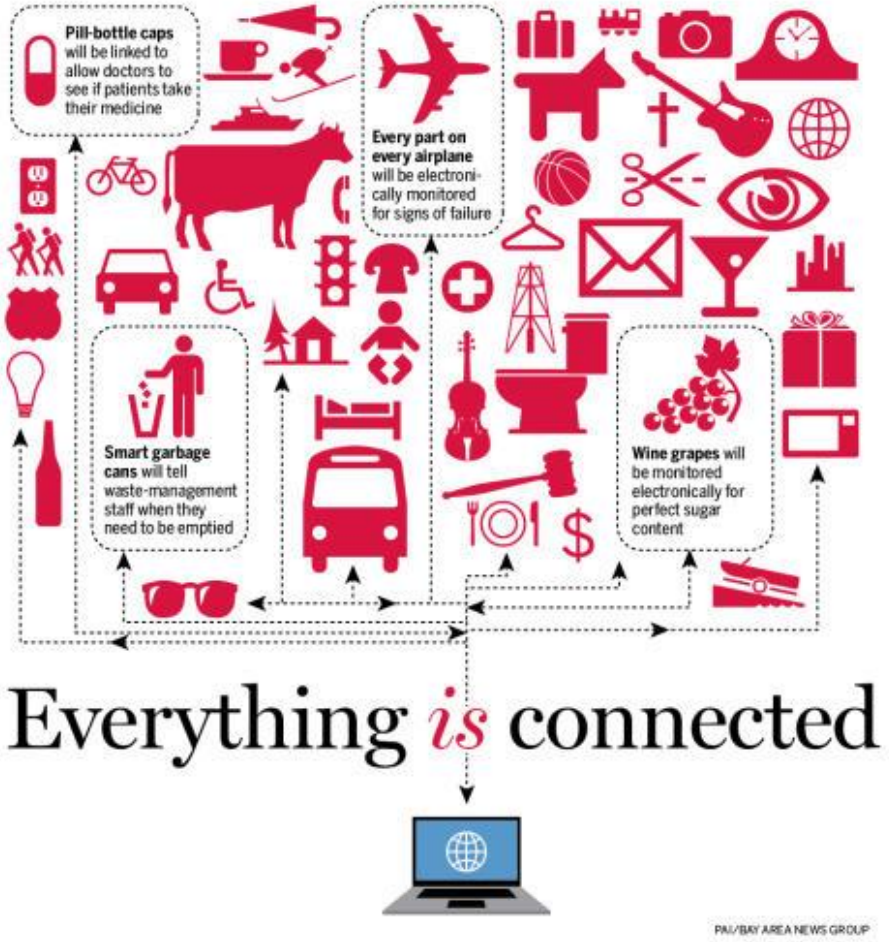




http://www.videcom.com/Portals/0/iphone1.png & http://images.techtimes.com/data/images/full/127370/events-for-gmail.jpg?w=600

Sistemas distribuidos de mayor tamaño…

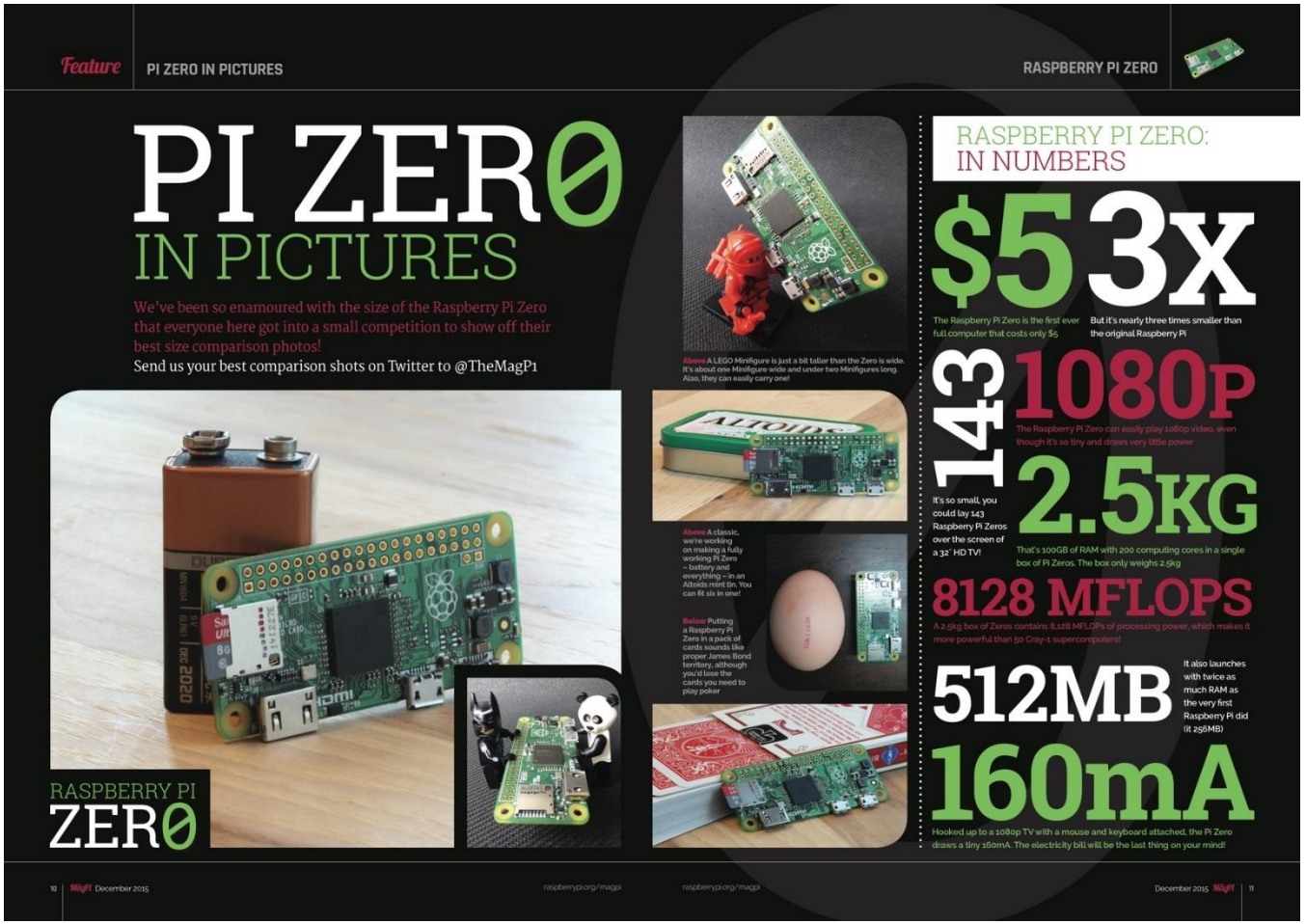
Todas las **cosas** conectadas



http://www.mercurynews.com/business/ci_24836116/internet-things-seen-bonanza-bay-area-businesses

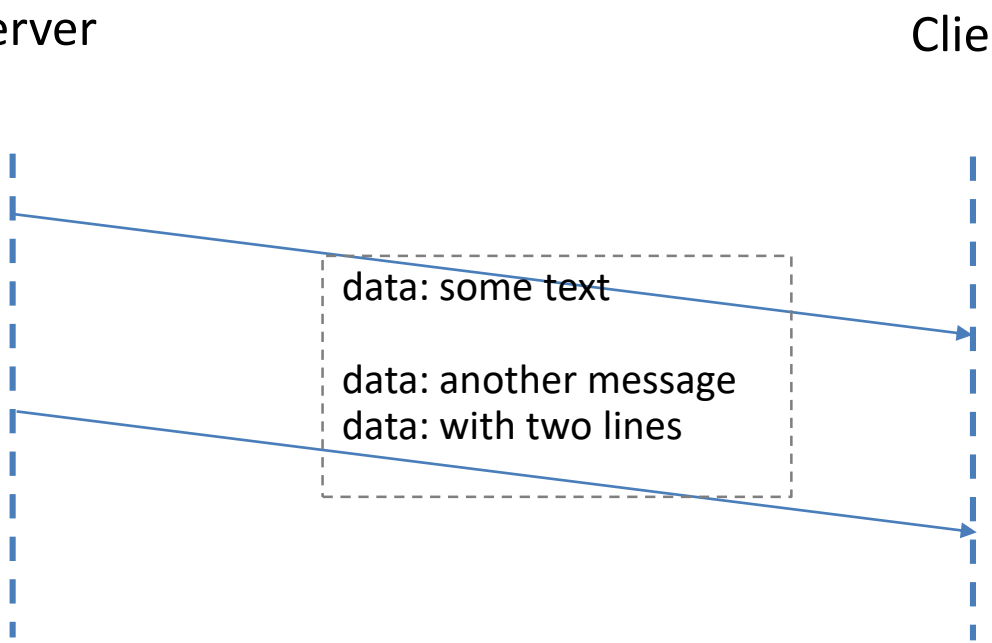
Ejemplo de IoT:

PI_0 + HTML5 Server-Sent Events



https://www.raspberrypi.org/magpi/wp-content/uploads/2015/11/Pi_Zero-Pics-Spread.jpg
https://redbear.cc/content/blog/pi-zero-iot-hat/

Ejemplo de IoT:
PI_0 + HTML5 Server-Sent Events



- Web page gets updates from a server
  - without ask for updates, the updates came automatically
- Examples: Twitter updates, stocks prices, news feeds, etc.

https://developer.mozilla.org/es/docs/Server-sent_events/utilizando_server_sent_events_sse

Ejemplo de IoT:

PI_0 + HTML5 Server-Sent Events

Server
(demo.php)

```
<?php
  header('Content-Type: text/event-stream');
  header('Cache-Control: no-cache');

  while (1)  {
    echo "data: {'time': " . date('r') . "}\n\n";
    ob_flush(); flush();
    sleep(1);
  }
?>
```

Client
(demo.html)

```
<span id="result"></span>
<script>
var source = new EventSource("demo.php");
source.onmessage = function (e) {
      var ref = document.getElementById("result") ;
      ref.innerHTML += e.data + "<br>";
};
source.onerror = function(e) { alert("Problems..."); };
</script>
```

- Web page gets updates from a server
  - without ask for updates, the updates came automatically
- Examples: Twitter updates, stocks prices, news feeds, etc.

http://www.w3schools.com/html/html5_serversentevents.asp

Ejemplo de IoT:

PI_0 + HTML5 Server-Sent Events

Server

(demo.sh)

```
echo "HTTP/1.1 200 OK"
echo "Access-Control-Allow-Origin: *"
echo "Content-Type: text/event-stream"
echo "Cache-Control: no-cache"
echo ""

while [ 1 ]; do
    T=$(date +%H:%M:%S)
    echo "data: {'timestamp': $T}\n\n"
    sleep 1
done
```

(demo-nc.sh)

```
./demo.sh | nc -l -p 8080
```

Client

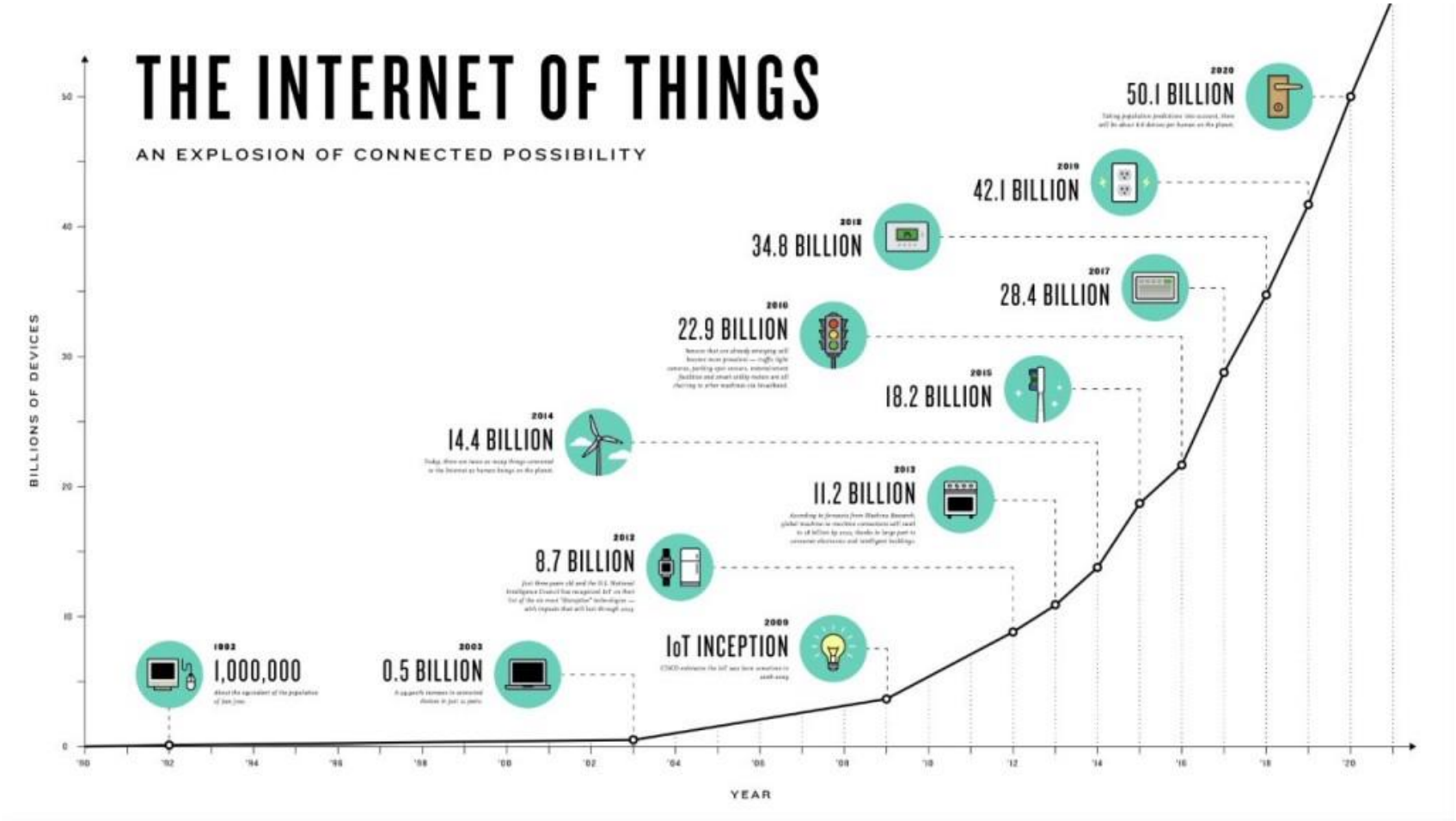
(demo.html)

```
<!DOCTYPE html>
<html>
<head>   <meta charset="utf-8" /> </head>
<body>
 <script>
   var s = new EventSource('http://<ip>:8080');
   s.onmessage = function(e) {
      document.body.innerHTML += e.data + '<br>';
   };
 </script>
</body>
</html>
```

https://github.com/acaldero/moon

Sistemas distribuidos de mayor tamaño…

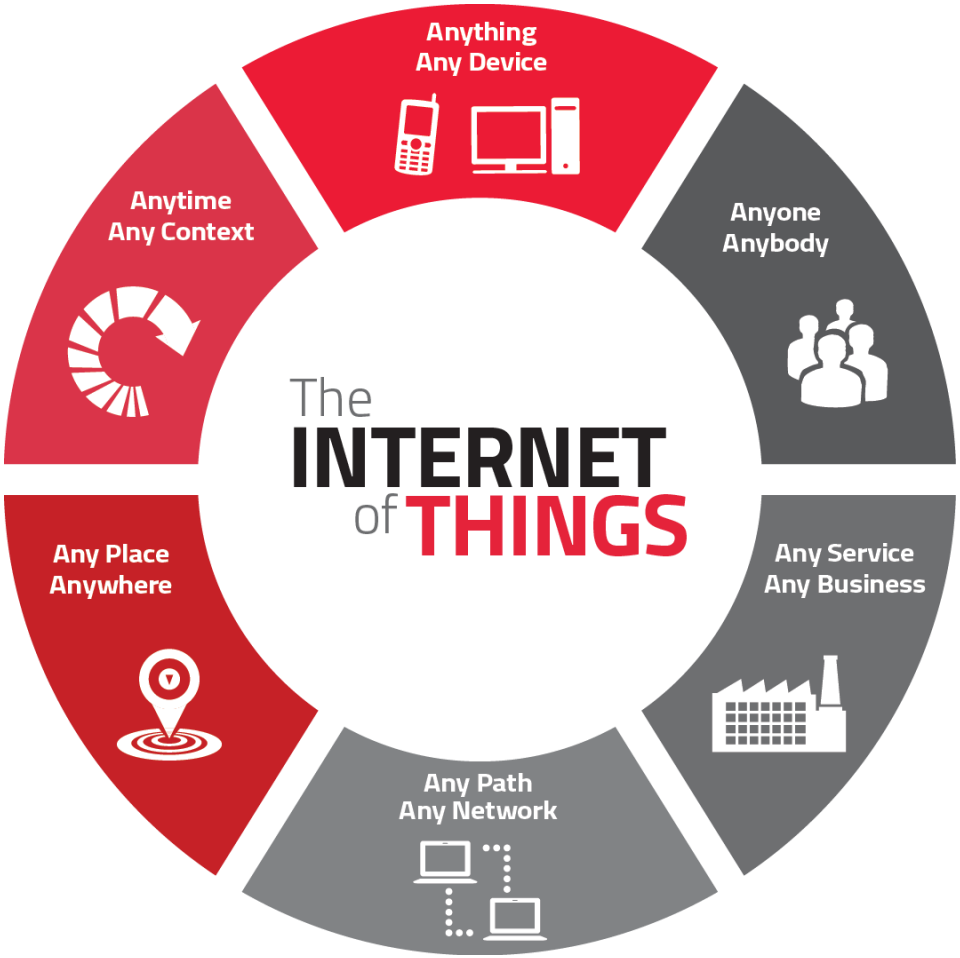
Internet of Things (IoT)



http://tarrysingh.com/2014/07/fog-computing-happens-when-big-data-analytics-marries-internet-of-things/

Sistemas distribuidos de mayor tamaño…

Internet of Things (IoT)



http://knowledgeblob.com/technology/a-brief-about-internet-of-things-iot/

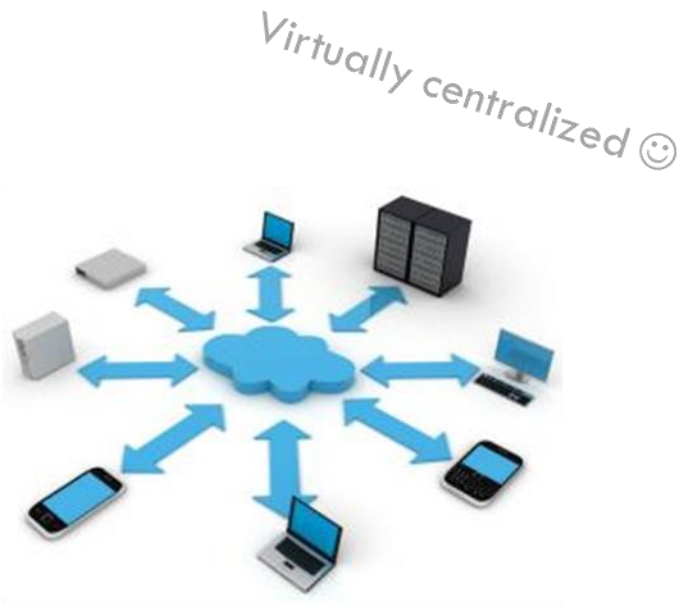
Contenidos

Big
Big
Big
Big
Big

- Inicio
- Transición
- (re)Evolución
- (re)Retos

http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838

Dependencia cada vez mayor…



Distributed

Distributed system:
existence of multiple elements is transparent

Cloud

Cloud computing:
Ubiquitous, on-demand access to shared pool of computing resources

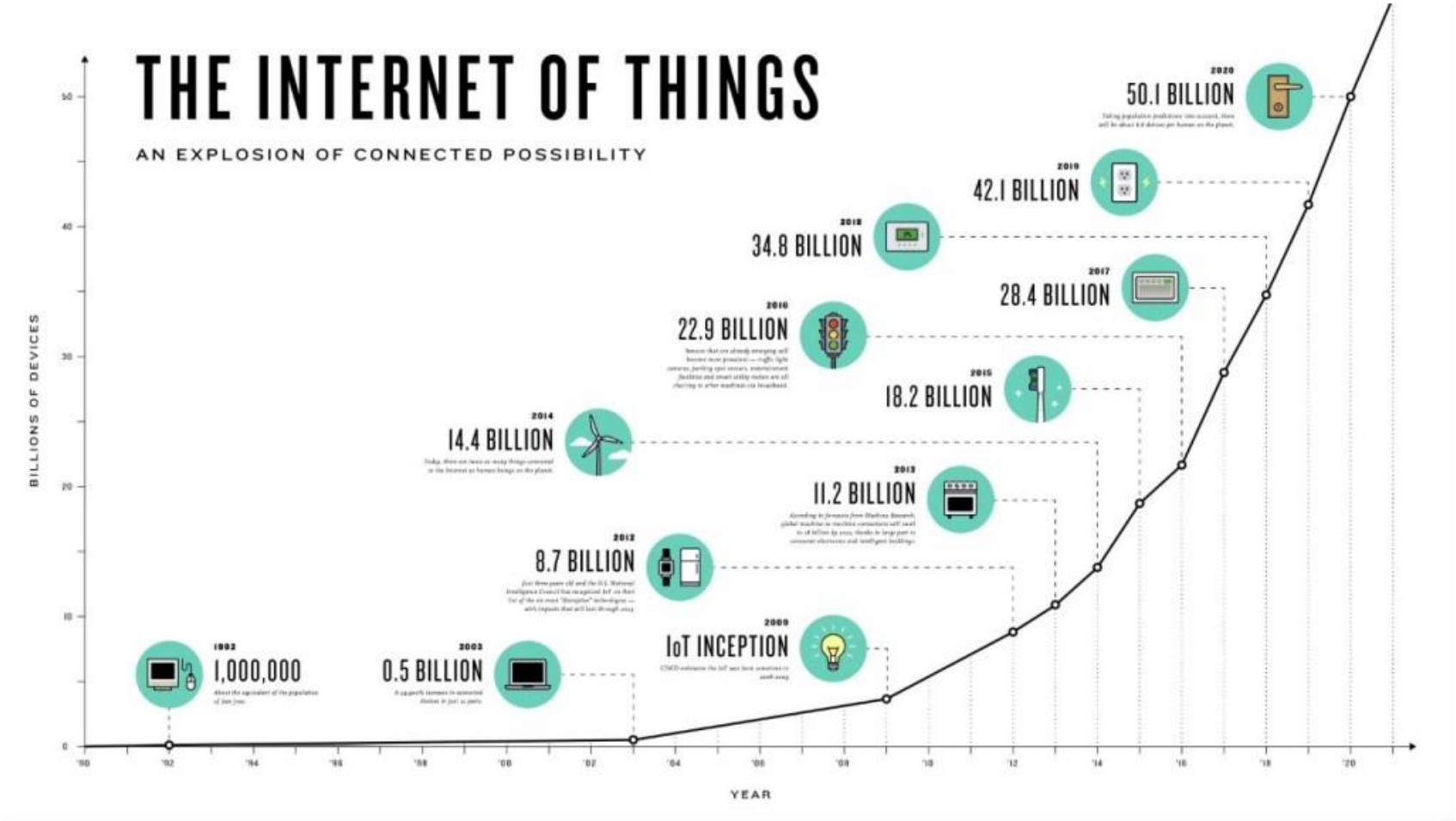
https://www.dcsorg.com/images/image_centralized_management.jpg
http://www.hsi.es/images/cc.png

Dependencia cada vez mayor…

criticidad con difícil vuelta atrás

http://kburnett.net/business-case/technology/mobility-2/

Nuevos retos…



http://tarrysingh.com/2014/07/fog-computing-happens-when-big-data-analytics-marries-internet-of-things/

Big data is coming...

From a Molehill To a Mountain

A timeline of digital data

Everyone knows that the amount of digital data has exploded in recent years. But it isn't always easy to put today's mountain of data in perspective.

To illustrate the data boom, here's a timeline with snapshots of some of the largest corporate data collections from each decade since the Univac computer was the hot new thing. —*Matthew Kassel*

2010s | **100 Petabytes**
Facebook Is Facebook's data hoard bigger than Google's today? Some say yes, some say no. According to Facebook, its user content makes up more than 100 petabytes of stored photos and video. And analyzing that data generates about 500 terabytes of new information every day—more than 2½ times the size of a '90s Wal-Mart data cache.

2000s | **25 Petabytes**
Google The explosion of the Internet in the '90s set the stage for Web-based companies to emerge in the following decade as the global leaders in big data.

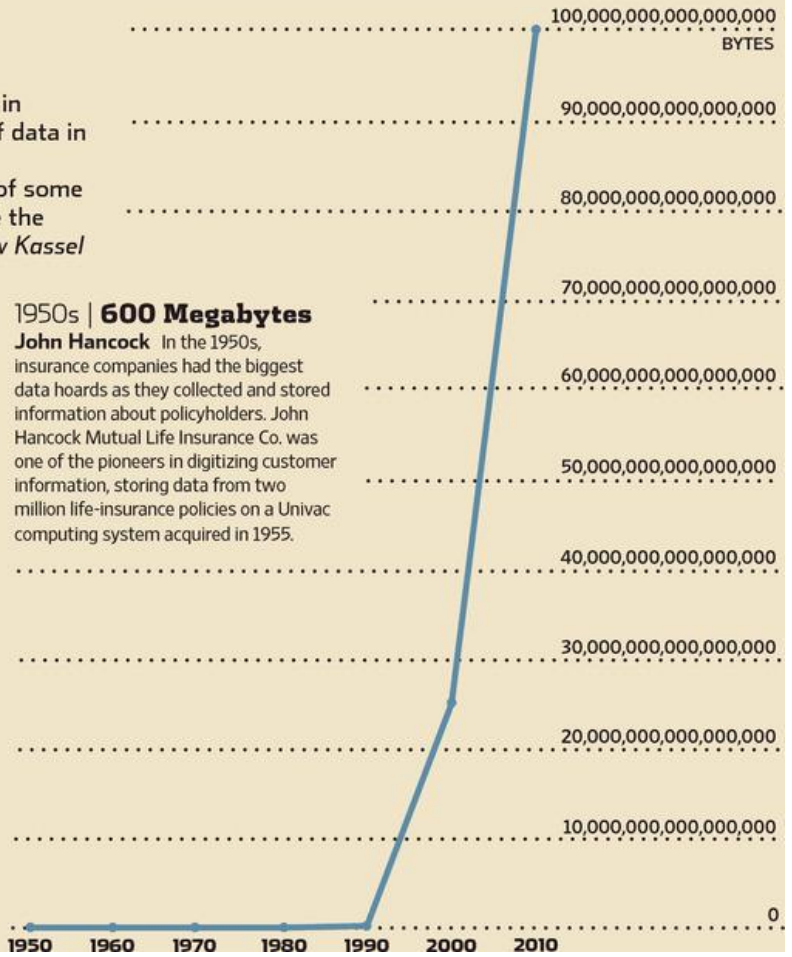
1990s | **180 Terabytes**
Wal Mart Wal-Mart in this decade became the largest American bricks-and-mortar retail operation, and, it is believed, had the biggest commercial data warehouse in the world. To put this number in perspective, experts have estimated that Amazon.com in the late '90s had single-digit terabytes of stored data.

1980s | **450 Gigabytes**
CitiCorp's NAIB In the 1980s, banks were at the forefront of data growth, with ATMs coming into vogue and banks focusing on collecting and analyzing transaction data for all their businesses.

1970s | **80 Gigabytes**
Federal Express Cosmos FedEx's Cosmos system, introduced in the 1970s, allowed the company to scan and track its huge volume of packages being shipped around the world.

1960s | **807 Megabytes**
American Airlines Sabre In the 1960s, American Airlines developed Sabre, a flight-reservation system built around one of the largest IBM computing systems available. Sabre was one of the first online computing systems, allowing the airline to keep track of an immense matrix of reservations, flight schedules and seat inventories.

1950s | **600 Megabytes**
John Hancock In the 1950s, insurance companies had the biggest data hoards as they collected and stored information about policyholders. John Hancock Mutual Life Insurance Co. was one of the pioneers in digitizing customer information, storing data from two million life-insurance policies on a Univac computing system acquired in 1955.



http://online.wsj.com/news/articles/SB10001424127887324178904578340071261396666

Big data is coming...

Ciencia
Industria
Logs de sistema
...

From a Molehill To a Mountain

A timeline of digital data

Everyone knows that the amount of digital data has exploded in recent years. But it isn't always easy to put today's mountain of data in perspective.

To illustrate the data boom, here's a timeline with snapshots of some of the largest corporate data collections from each decade since the Univac computer was the hot new thing. —*Matthew Kassel*

2010s | **100 Petabytes**
Facebook Is Facebook's data hoard bigger than Google's today? Some say yes, some say no. According to Facebook, its user content makes up more than 100 petabytes of stored photos and video. And analyzing that data generates about 500 terabytes of new information every day—more than 2½ times the size of a '90s Wal-Mart data cache.

2000s | **25 Petabytes**
Google The explosion of the Internet in the '90s set the stage for Web-based companies to emerge in the following decade as the global leaders in big data.

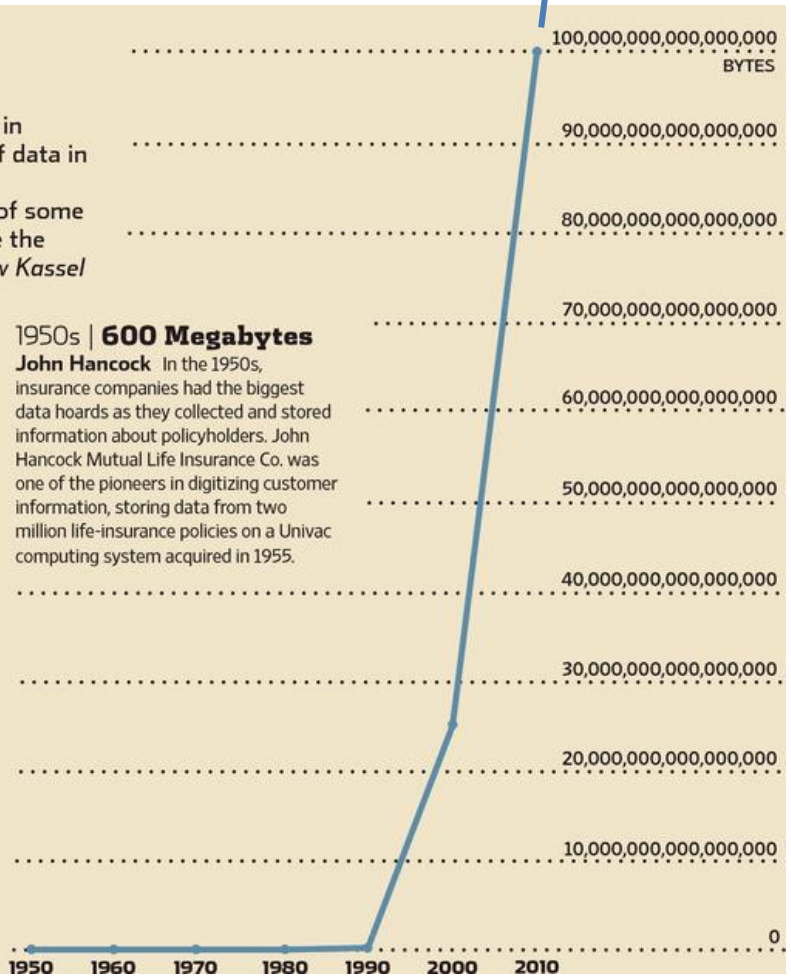
1990s | **180 Terabytes**
Wal Mart Wal-Mart in this decade became the largest American bricks-and-mortar retail operation, and, it is believed, had the biggest commercial data warehouse in the world. To put this number in perspective, experts have estimated that Amazon.com in the late '90s had single-digit terabytes of stored data.

1980s | **450 Gigabytes**
CitiCorp's NAIB In the 1980s, banks were at the forefront of data growth, with ATMs coming into vogue and banks focusing on collecting and analyzing transaction data for all their businesses.

1970s | **80 Gigabytes**
Federal Express Cosmos FedEx's Cosmos system, introduced in the 1970s, allowed the company to scan and track its huge volume of packages being shipped around the world.

1960s | **807 Megabytes**
American Airlines Sabre In the 1960s, American Airlines developed Sabre, a flight-reservation system built around one of the largest IBM computing systems available. Sabre was one of the first online computing systems, allowing the airline to keep track of an immense matrix of reservations, flight schedules and seat inventories.

1950s | **600 Megabytes**
John Hancock In the 1950s, insurance companies had the biggest data hoards as they collected and stored information about policyholders. John Hancock Mutual Life Insurance Co. was one of the pioneers in digitizing customer information, storing data from two million life-insurance policies on a Univac computing system acquired in 1955.



http://online.wsj.com/news/articles/SB10001424127887324178904578340071261396666

Big Data frequently used...

- Text mining
- Index building
- Graph creation and analysis
- Pattern recognition
- Prediction model
- …

http://www.ndm.net/emcstore/storage/greenplum
http://fcw.com/~/media/GIG/FCWNow/Topics/Big%20Data/Big_data.png

Big Data & Big Processing…

https://www.linkedin.com/pulse/open-source-network-has-organized-workshop-big-data-unlock-sharmin

Las transiciones siguen sin ser fáciles…

Áreas de trabajo…

- Heterogeneity
- Openness
- Security
- Scalability
- Failure Handling
- Concurrency
- Transparency

- A system is scalable if will remain effective when there is a significant increase in the number of resources and the number of users (represented by client programs).

Big Distributed Systems are coming...

- Heterogeneity
- Openness
- Security
- Scalability
- Failure Handling
- Concurrency
- Transparency

- A system is scalable if will remain effective when there is a significant increase in the number of resources and the number of users (represented by client programs).
- Many challenges:
  - Control of performance degradation
  - Cost of physical resources
    - But preventing running out of resources
    - But avoiding performance bottlenecks

Contenidos

Big
Big
Big
Big
Big

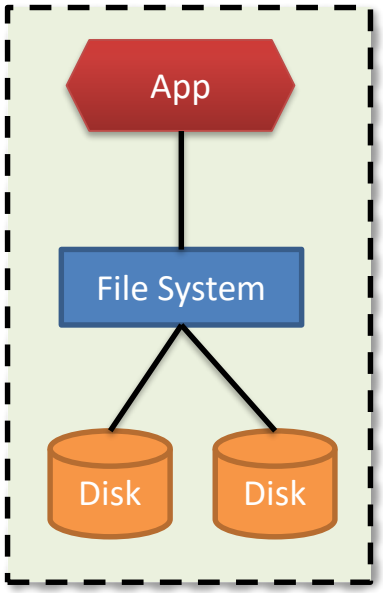
- Inicio
- Transición
- (re)Evolución
- (re)Retos

- **Inicio**
- Hadoop
- Ecosistema
- Evolución

http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838
http://gruposda.es/wp-content/uploads/sertecni.jpg

Repaso de opciones… (1/3)

DAS
Direct Attached Storage

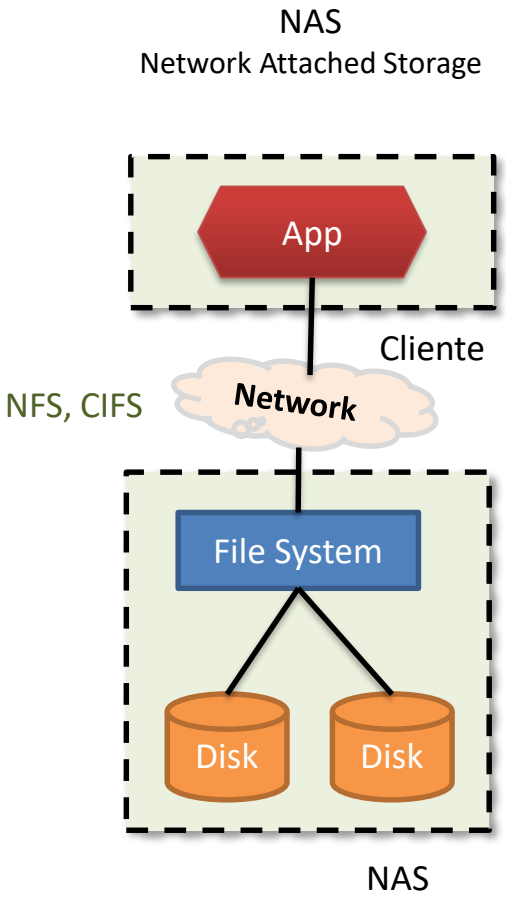


Ordenador

- V/I:
  - Rapidez, simplicidad
  - No compartición, crecimiento limitado

http://www.scsi4me.com/images/1296909213_60.jpg

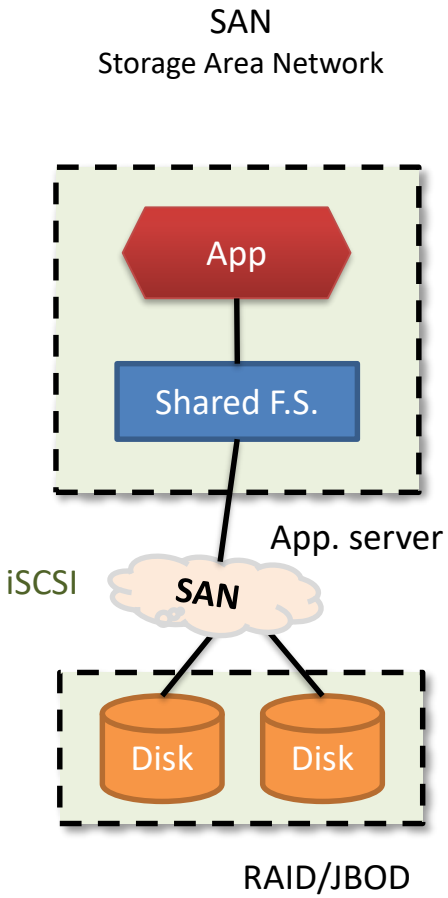
Repaso de opciones… (2/3)

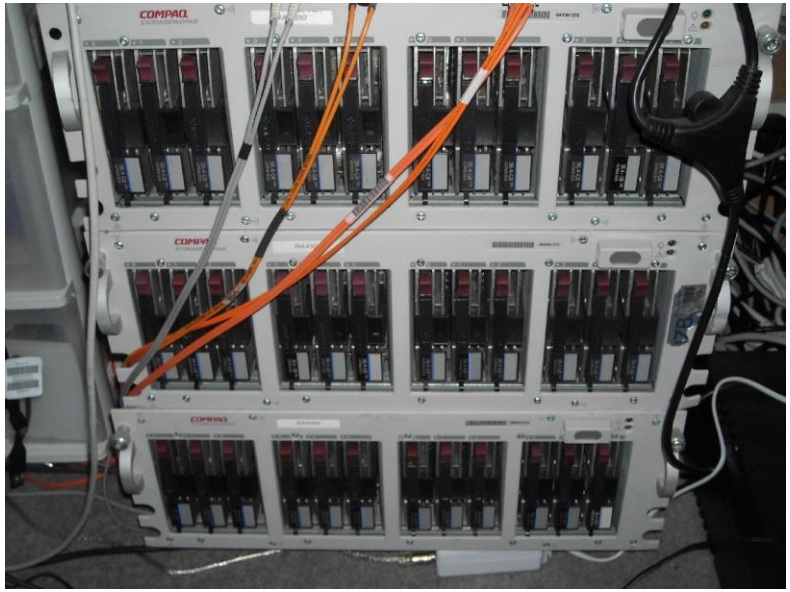


- V/I:
  - Compartición (RO), crecimiento
  - Red limita velocidad y crecimiento

http://www.jollynas.com/nas/img/wJollyNAS.jpg

Repaso de opciones… (3/3)



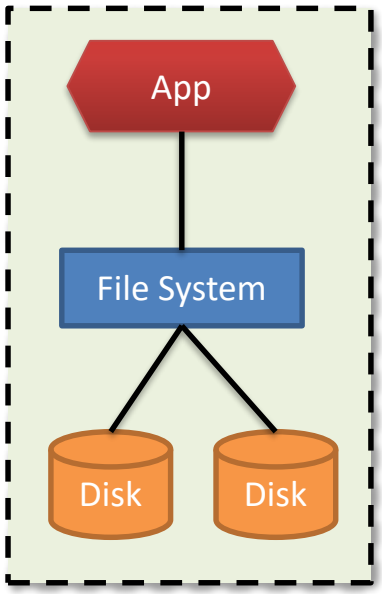


- V/I:
  - Mejor crecimiento-velocidad
  - Complejidad, coste

http://andysworld.org.uk/blog/wp-content/uploads/2010/05/ra4100.jpg

Combinación de opciones...

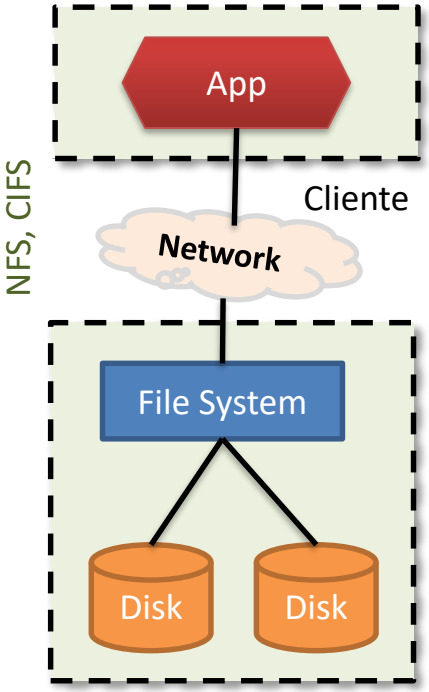
DAS
Direct Attached Storage



Ordenador

+

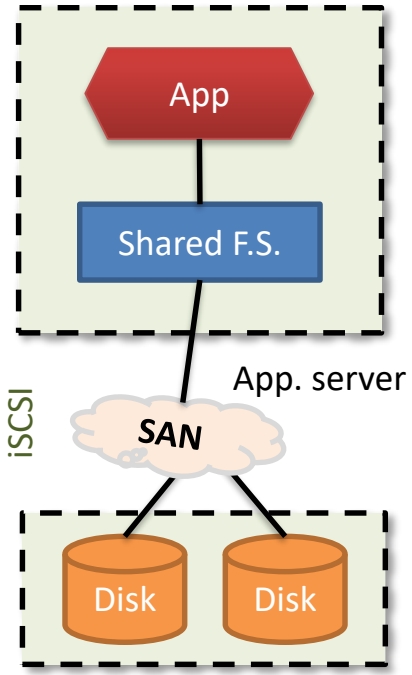
NAS
Network Attached Storage



NAS

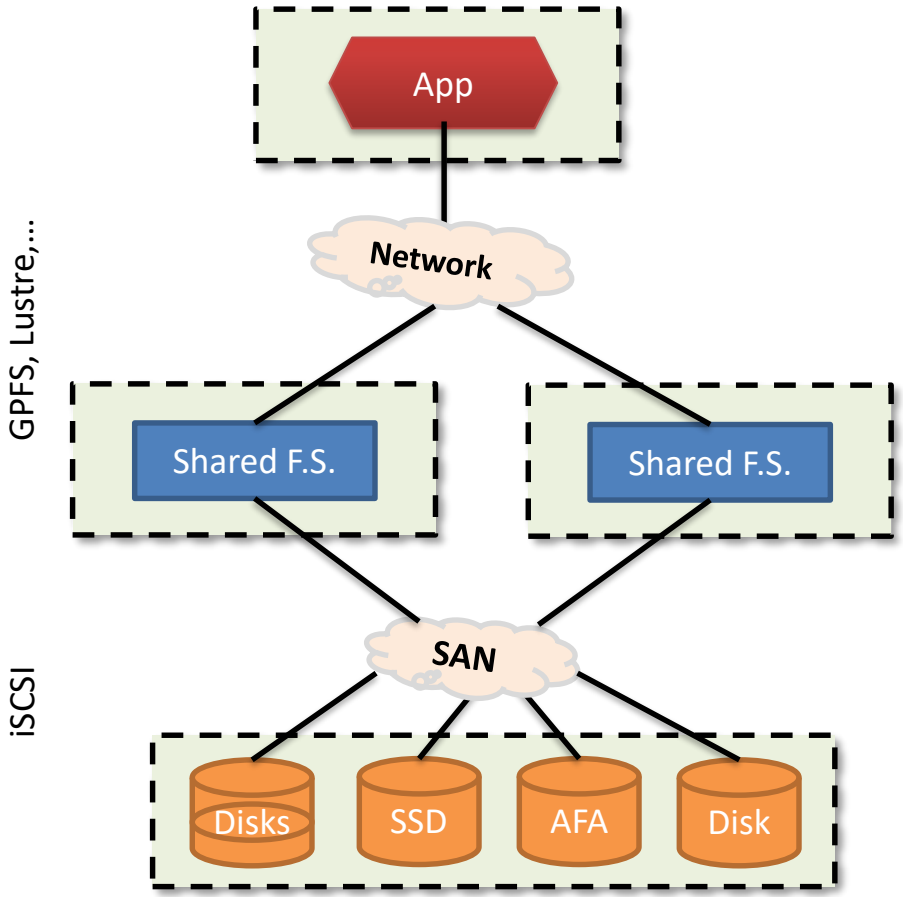
+

SAN
Storage Area Network



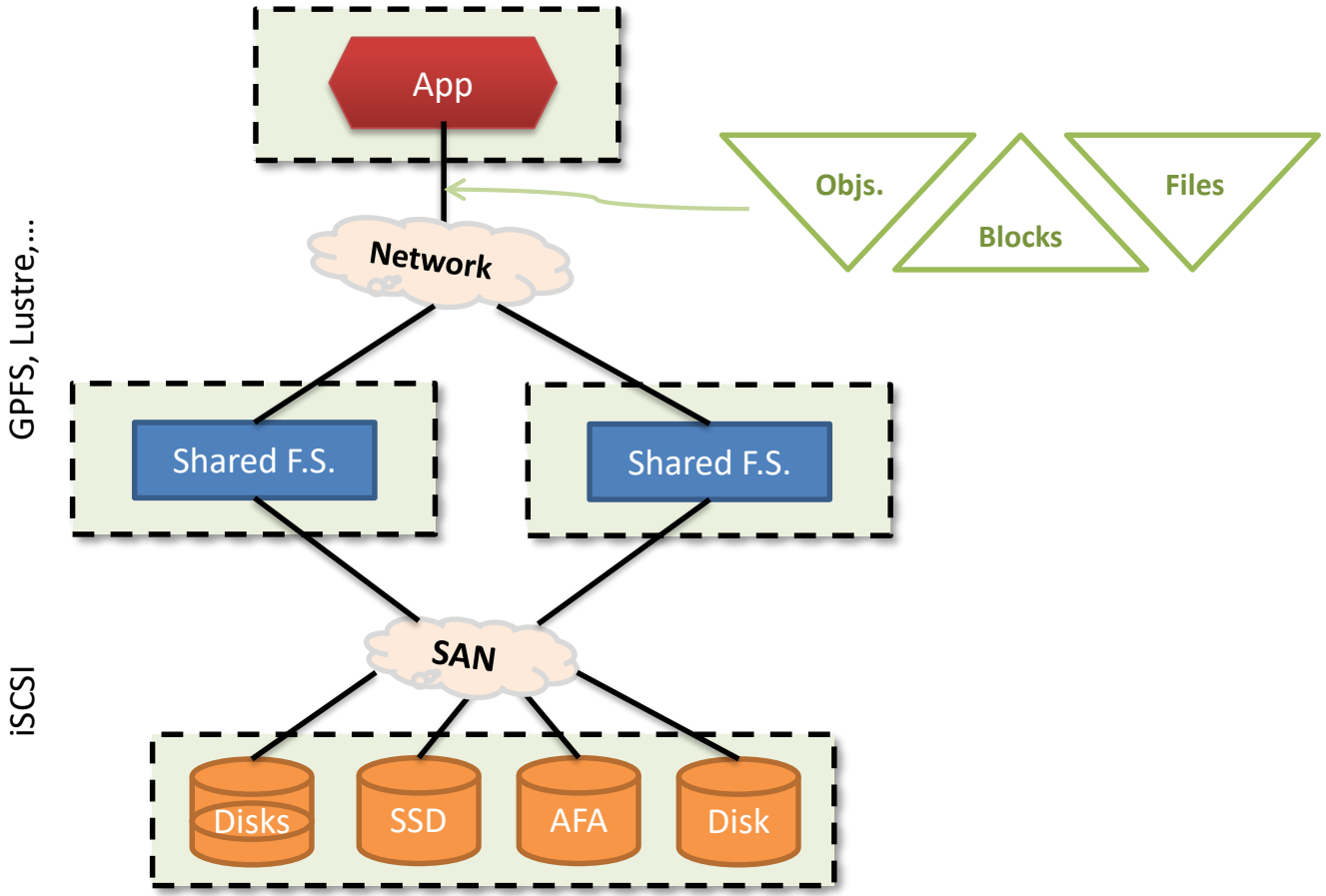
RAID/JBOD

Combinación de opciones...

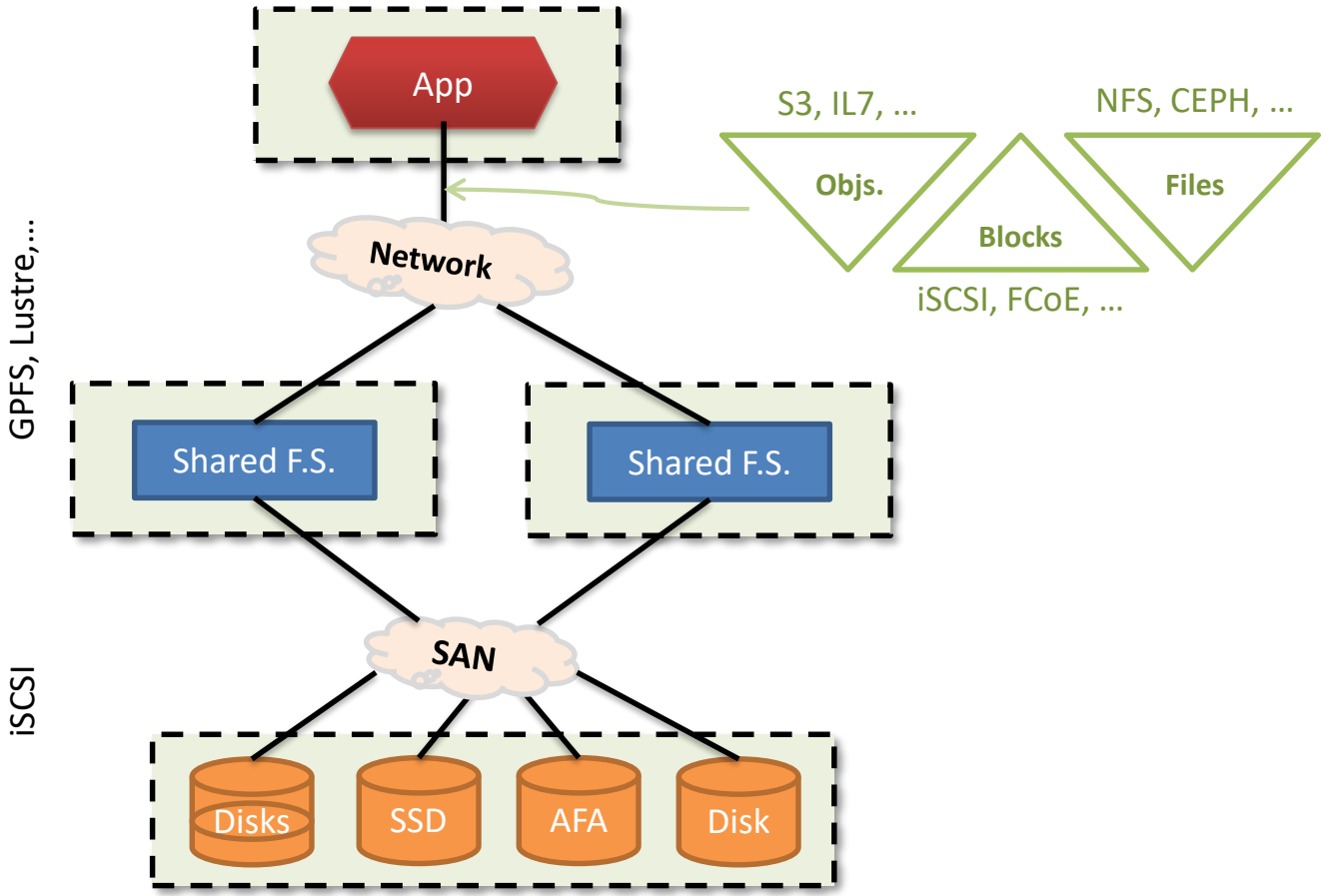


NAS + SAN

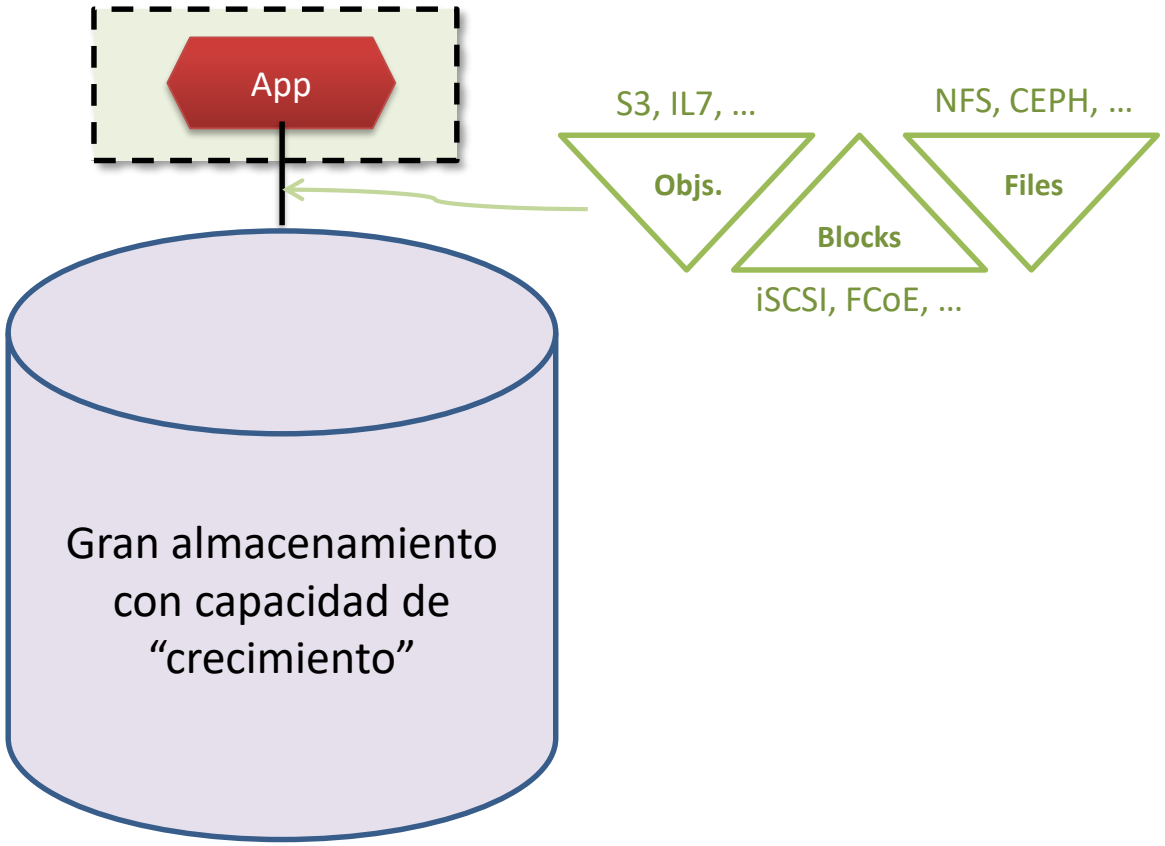
Combinación de opciones...



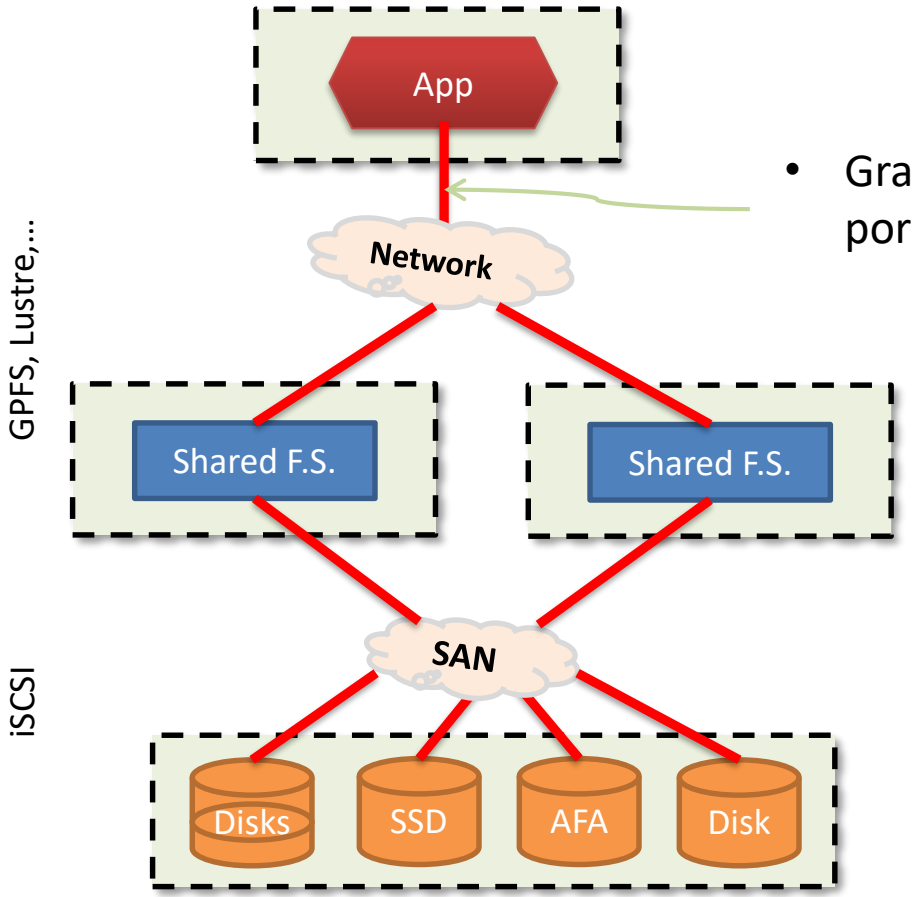
Combinación de opciones...



Combinación de opciones...

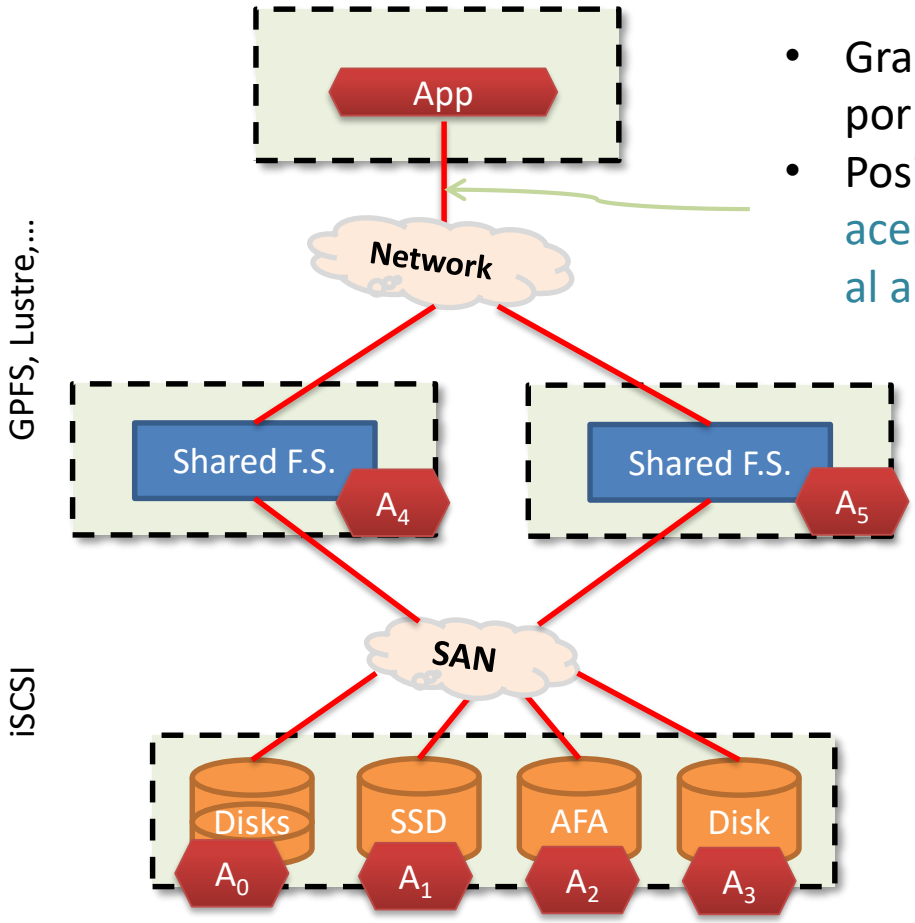


Problemas…



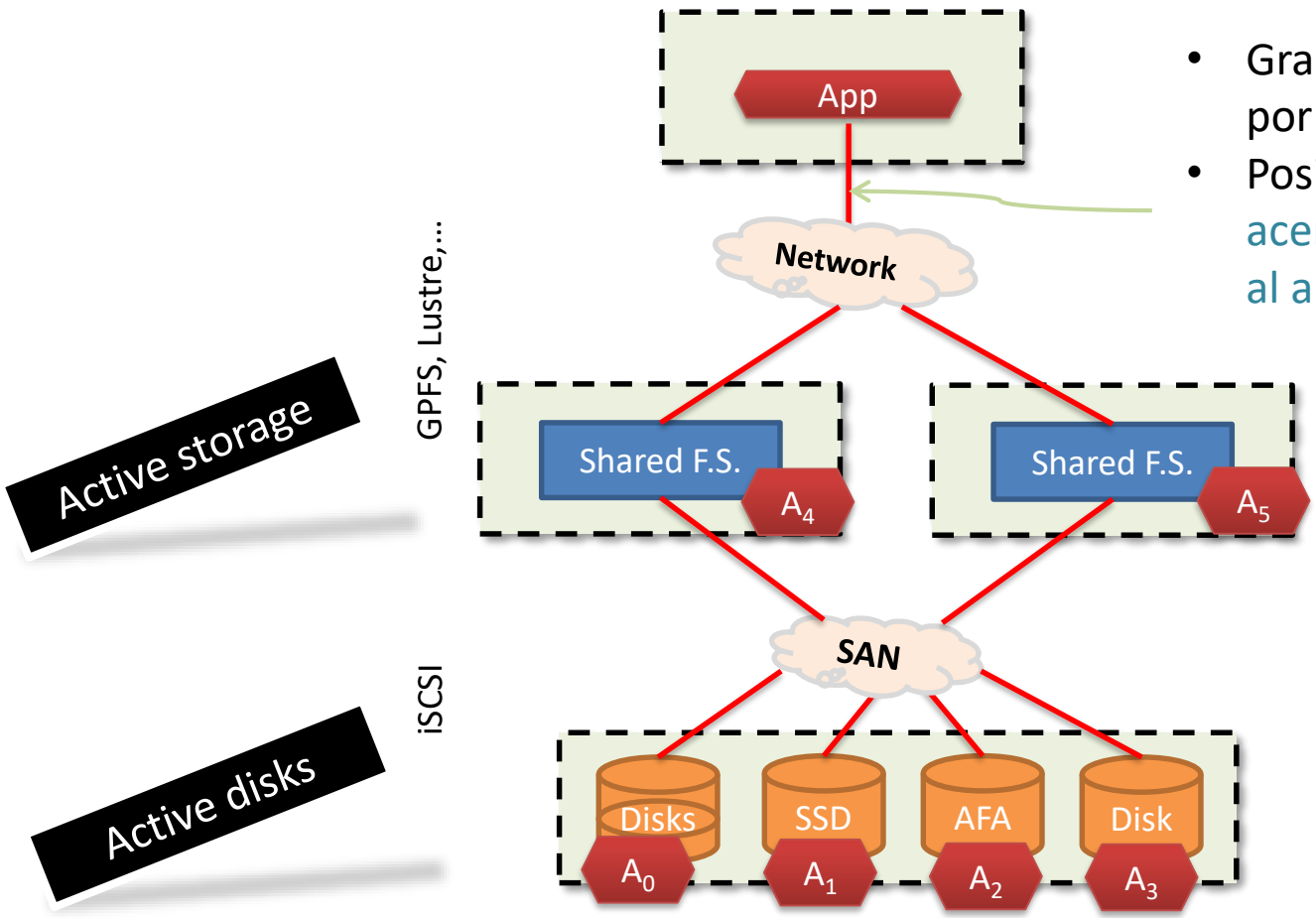
- Gran cantidad de tráfico por movimiento de datos.

Opciones para almacenamiento…



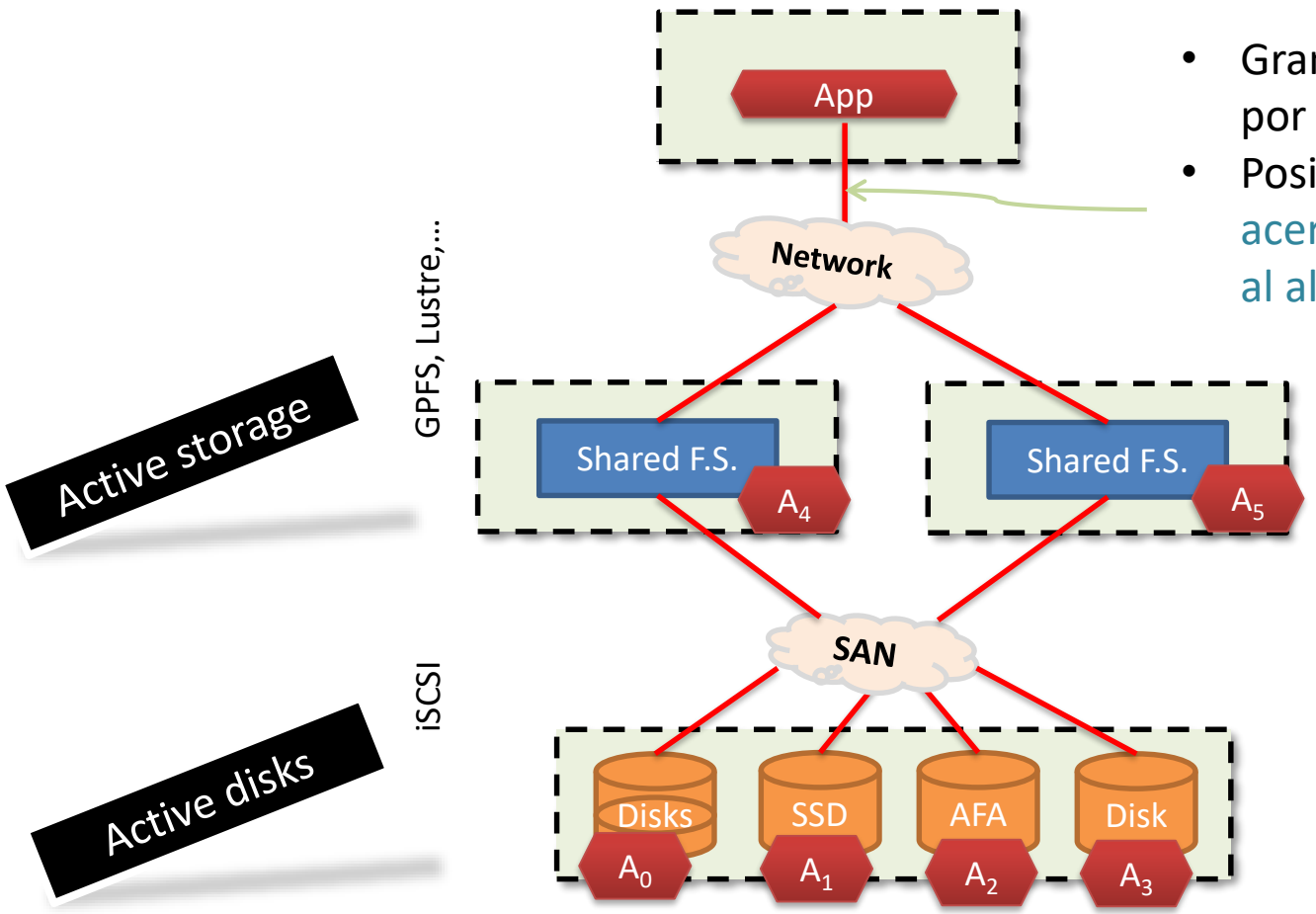
- Gran cantidad de tráfico por movimiento de datos.
- Posible disminuir si se acerca parte del cómputo al almacenamiento.

Opciones para almacenamiento…



- Gran cantidad de tráfico por movimiento de datos.
- Posible disminuir si se acerca parte del cómputo al almacenamiento.

Opciones para almacenamiento…



- Gran cantidad de tráfico por movimiento de datos.
- Posible disminuir si se acerca parte del cómputo al almacenamiento.

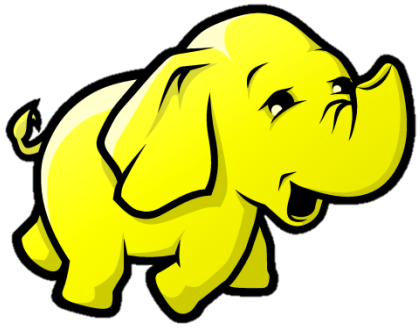
http://rsoltd.com/wp-content/uploads/2014/04/expensive.jpg

Buscar/crear la herramienta adecuada…

http://storageio.com/images/SIO_ToolBox.png

Contenidos

Big
Big
Big
Big
Big



- Inicio
- Transición
- (re)Evolución
- (re)Retos

- Inicio
- **Hadoop**
- Ecosistema
- Evolución

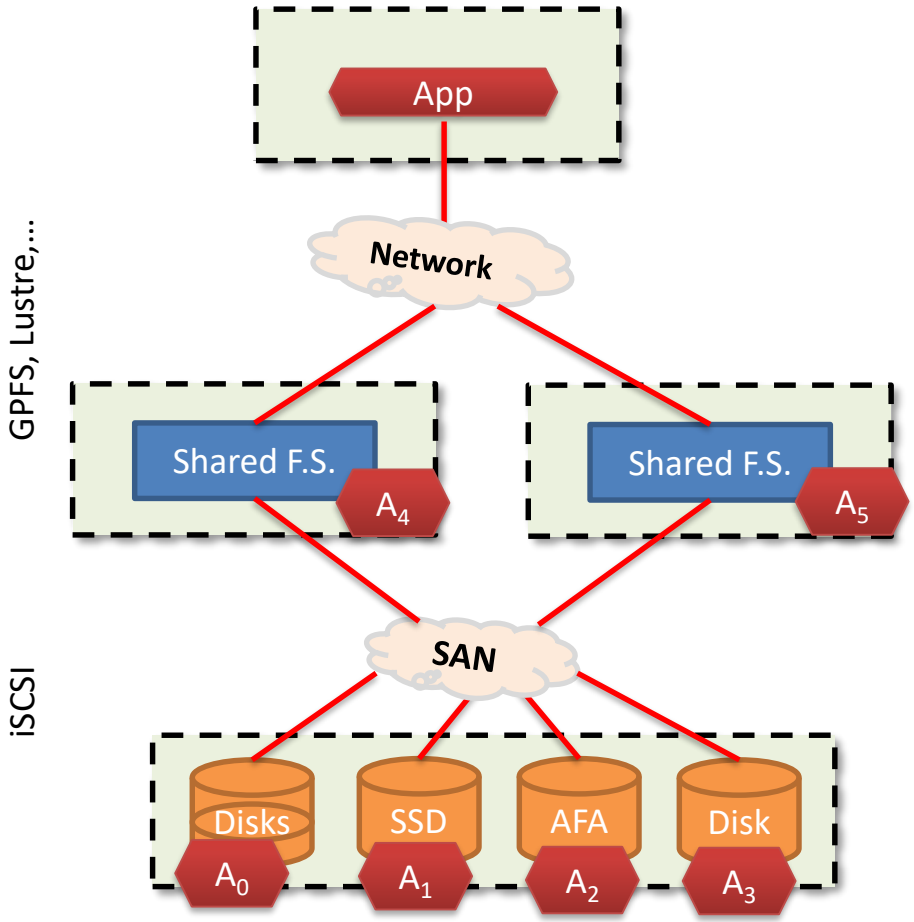
http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838
http://gruposda.es/wp-content/uploads/sertecni.jpg

The Google File System

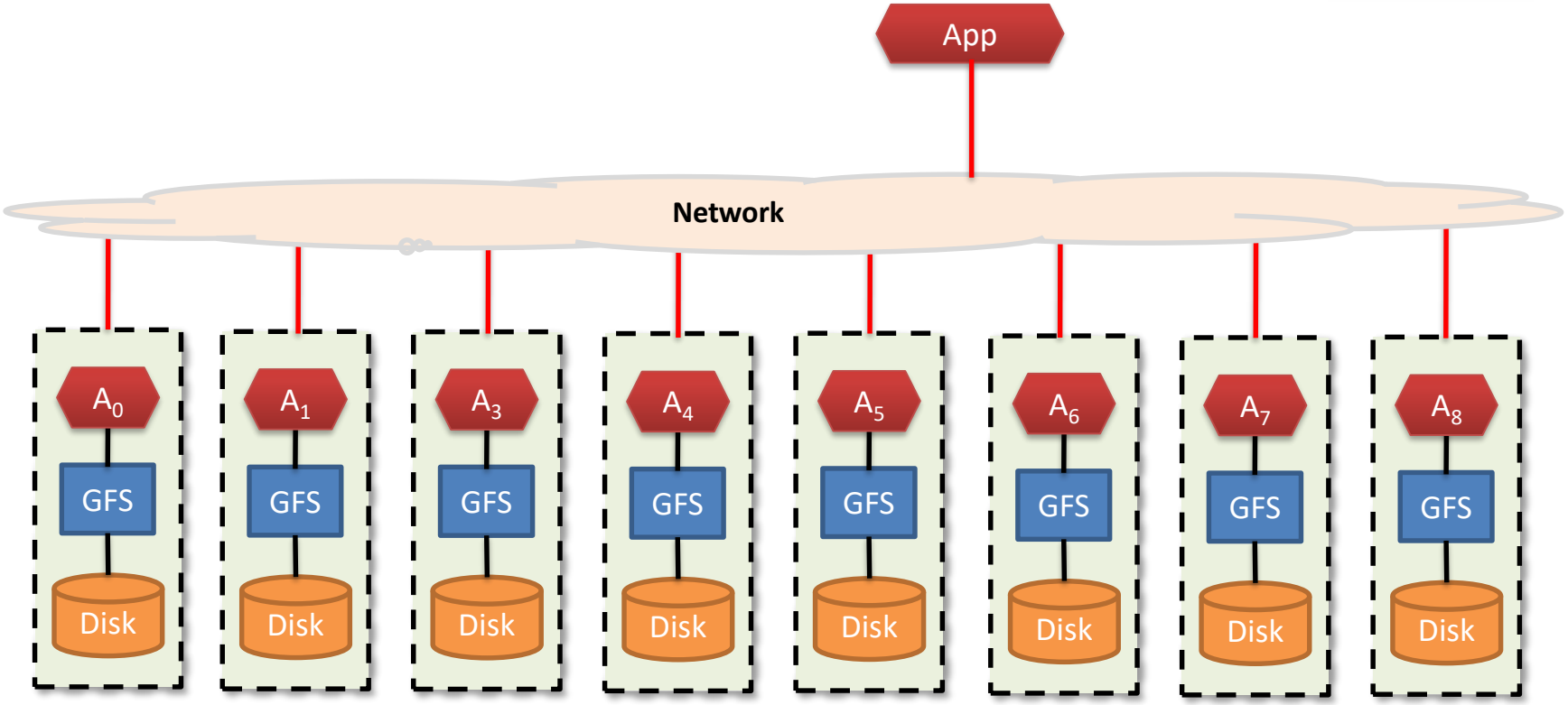
- Google presenta MapReduce y Google File System (GFS)
  - MapReduce es una propuesta para aplicar la misma función a particiones de datos (map) y luego se tiene el resultado procesando los resultados parciales (reduce)
  - GFS es una propuesta para almacenar petabytes de datos en muchas máquinas comunes, tratando con fallos, distribución, etc.

http://static.googleusercontent.com/media/research.google.com/en//archive/gfs-sosp2003.pdf

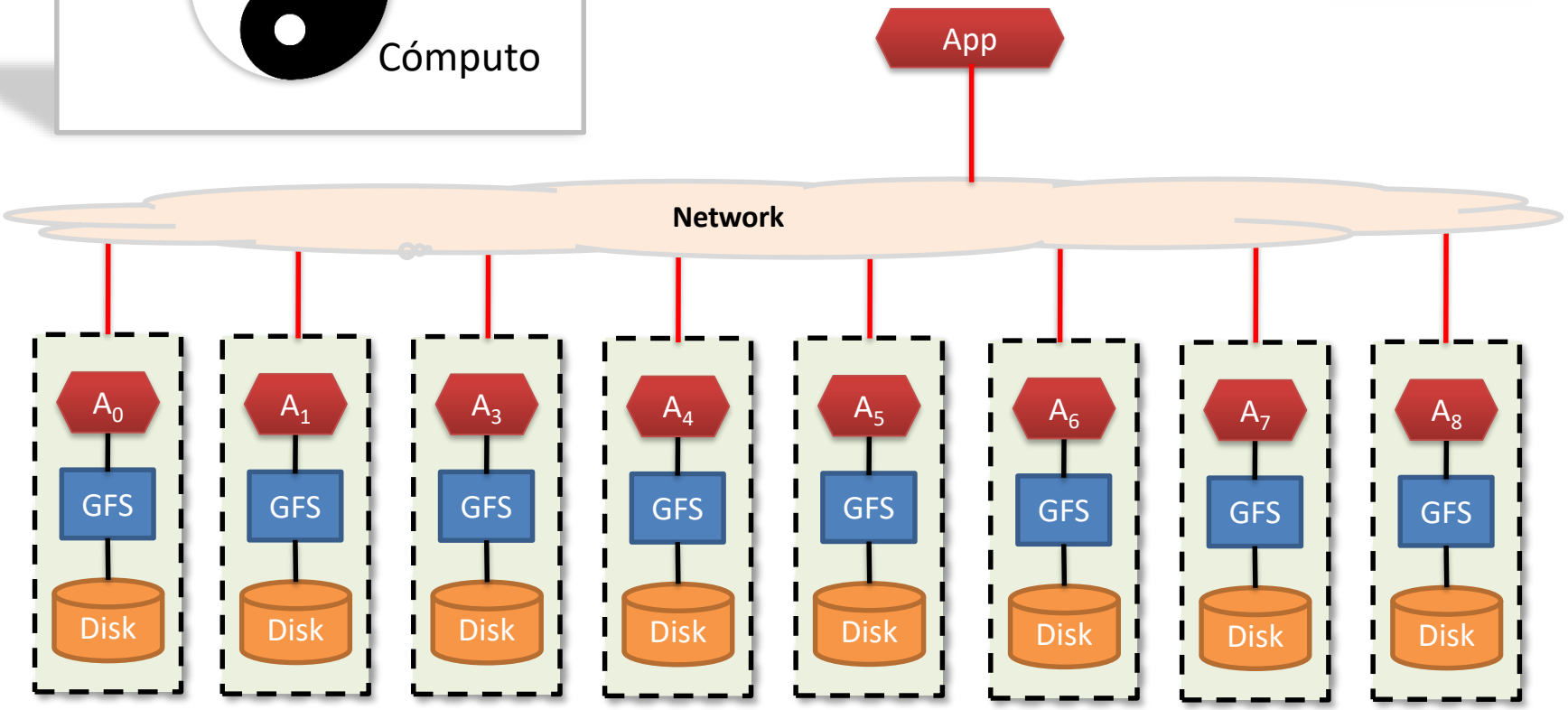
De esta opción…



...a esta opción



...a esta opción



https://upload.wikimedia.org/wikipedia/commons/thumb/7/7c/Yin_and_Yang.svg/1024px-Yin_and_Yang.svg.png

Breve historia

- Doug Cutting trabajando en Yahoo!
  e inspirado por estas tecnologías,
  inicia el desarrollo de **Hadoop**.
  - El proyecto usa Java, implanta las ideas detrás de GFS y MapReduce.
  - El logotipo se basa en el elefante amarillo que era el juguete favorito de su hijo.
  - Doug Cutting pasó a trabajar a cloudera.

http://www.enriquedans.com/2011/11/hadoop-el-elefante-omnipresente.html

Breve historia

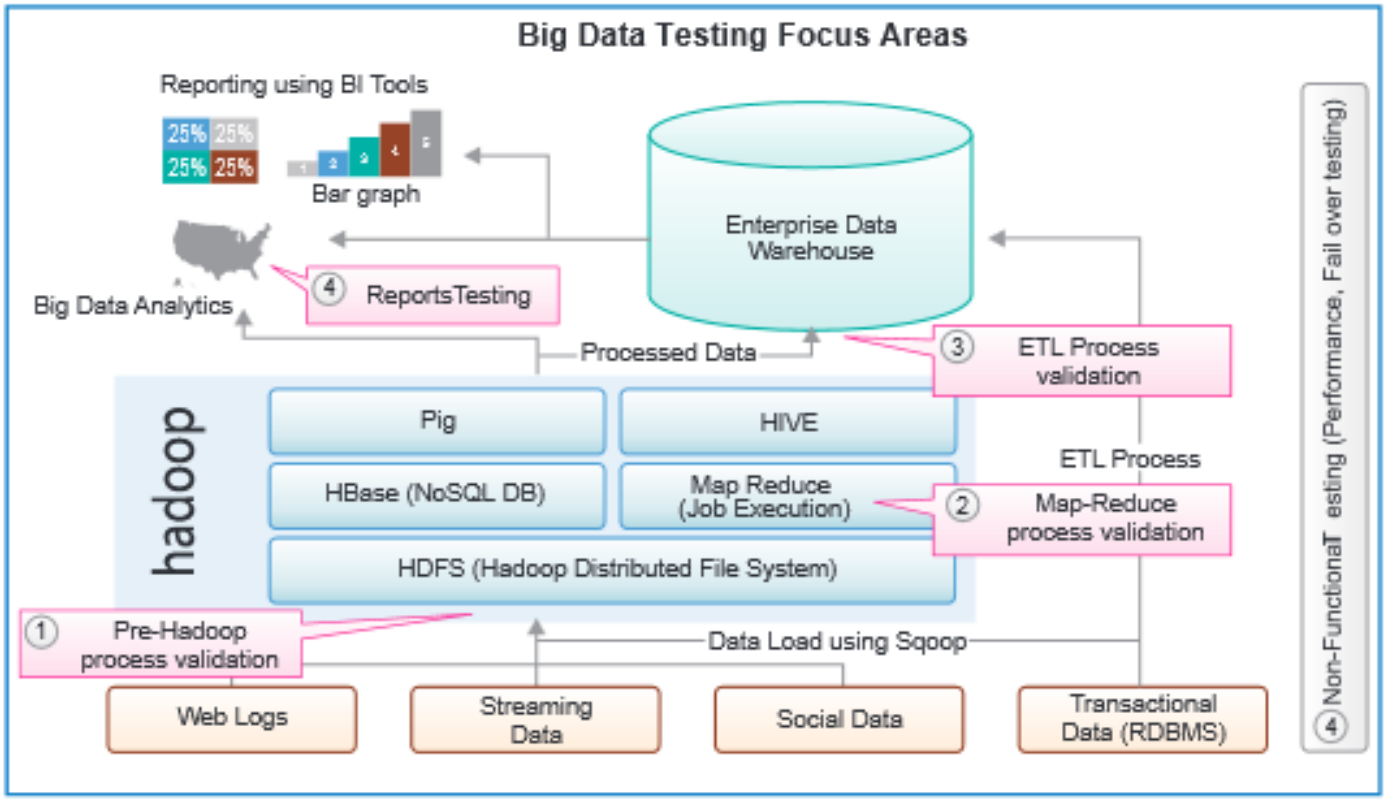
- Actualmente es un proyecto de código abierto bajo licencia Apache.
  - Su licencia ha facilitado que sea adoptado por un importante número de empresas.
- El apoyo de IBM, Oracle, EMC, etc. ha acelerado su implantación y su mejora en prestaciones.
  - Gran uso en proyectos tipo *big data*.
  - JPMorgan Chase: “We’re hiring, and we’re paying 10% more than the other guys.”

http://www.informationweek.com/software/information-management/its-next-hot-job-hadoop-guru/d/d-id/1101209?

http://hadoop.apache.org

- Apache Hadoop es un proyecto software *open source* para computación distribuida escalable y de confianza (*reliable*).
- Ofrece un framework que permite la computación distribuida de grandes conjuntos de datos mediante clusters de ordenadores usando modelos de programación simples.
  - Diseñado para poder pasar de un solo servidor a miles de máquinas (scale-up), donde cada una ofrece tanto computación como almacenamiento.
  - El framework está diseñado para detectar y tratar fallos a nivel de aplicación. Esto permite ofrecer servicios con alta disponibilidad sobre un cluster de ordenadores (aunque estos tengan fallos).

B.D.A. Workflow example…



http://www.infoivy.com/2013/12/5-steps-for-big-data-application.html

Big Opportunities…

Big Opportunities

A study by International Data Corp. cited data from these sources as being particularly ripe for analysis:

SURVEILLANCE FOOTAGE

There is greater opportunity to embed more intelligence into digital surveillance cameras so that footage can be captured, analyzed and tagged in real time. This can expedite crime investigations, enhance retailers' analysis of consumer traffic patterns and improve military intelligence.

MEDICAL DEVICES

Various kinds of sensors, including some that may be implanted in the body, have the potential to capture biometrics, track the effectiveness of medicines, correlate bodily activity with health, monitor potential outbreaks of viruses and more, all in real time.

SOCIAL MEDIA

Information freely published by massive groups of people through online social outlets can be a great source of data to help bring the next big thing to market, predict the outcome of elections, and much more.

CONSUMER IMAGES

People say a lot about themselves when they post pictures of themselves or family and friends online. The key will be the introduction of tagging algorithms that can analyze images in real time when pictures are uploaded, or en masse after they are aggregated from various sites.

Source: International Data Corp.

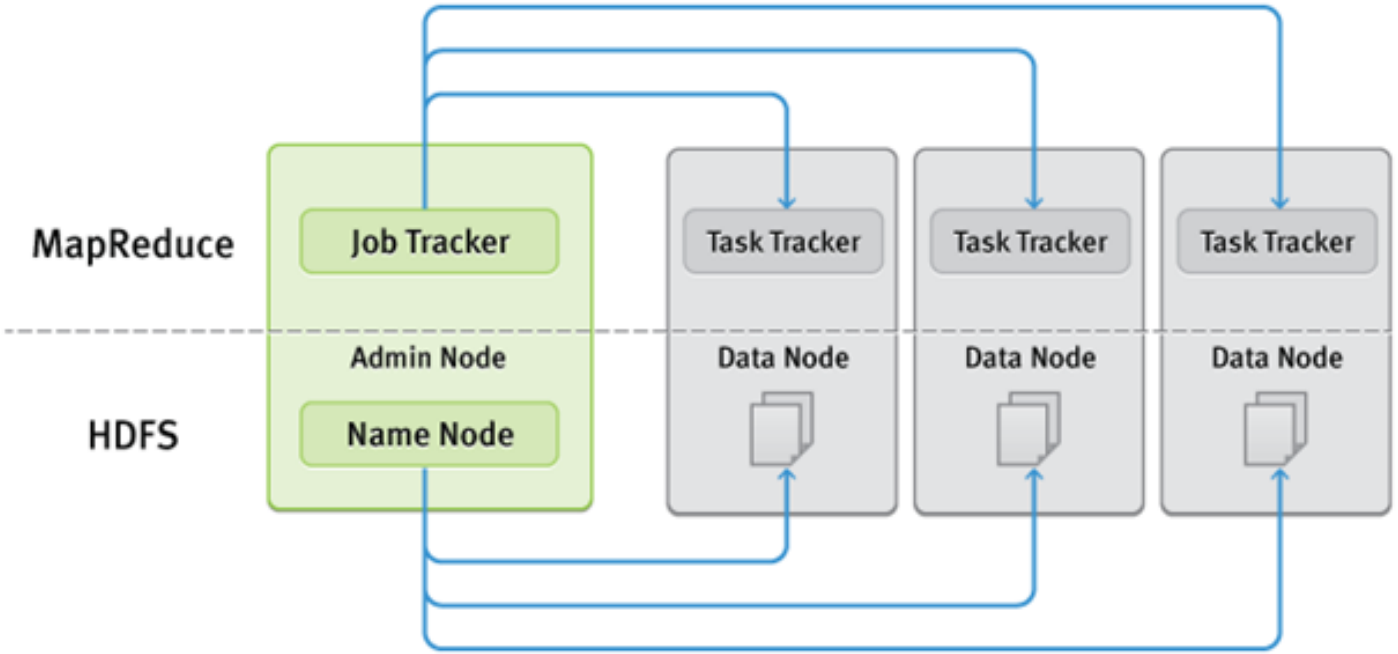
The Wall Street Journal

http://online.wsj.com/news/articles/SB10001424127887324178904578340071261396666

Big Opportunities…

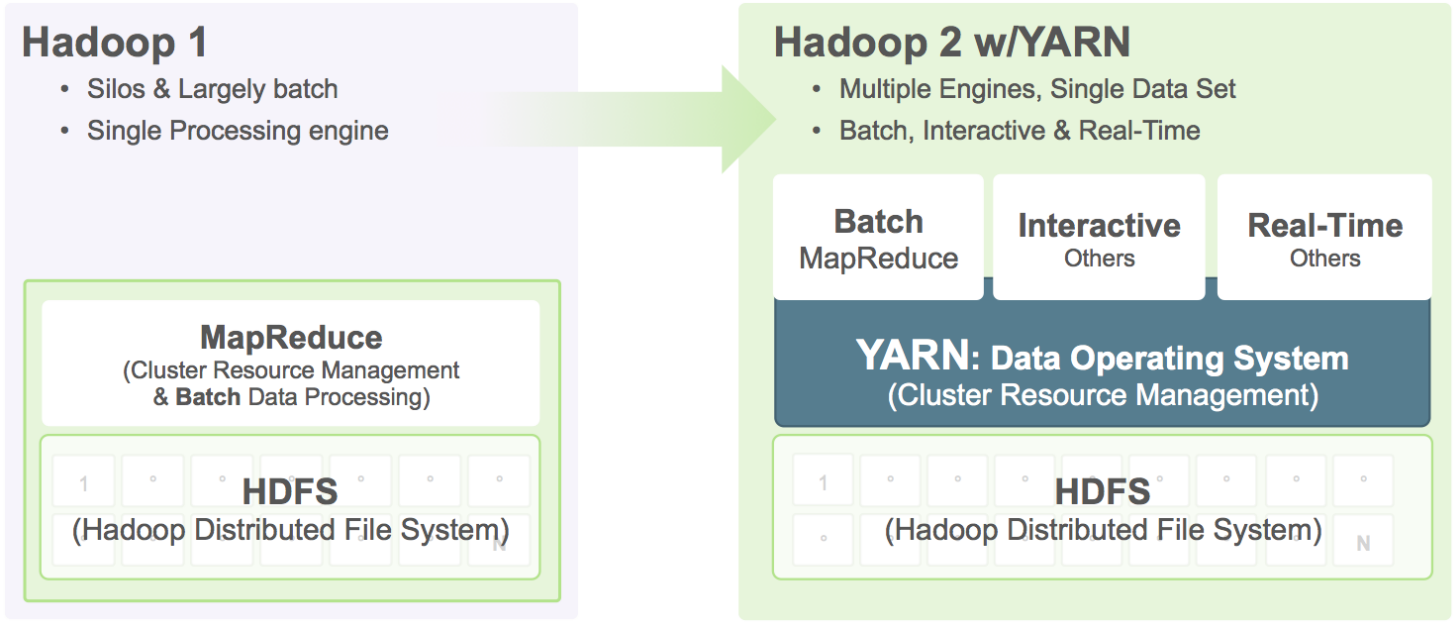
http://www.indeed.com/jobtrends?q=Big-data&relative=1

Arquitectura



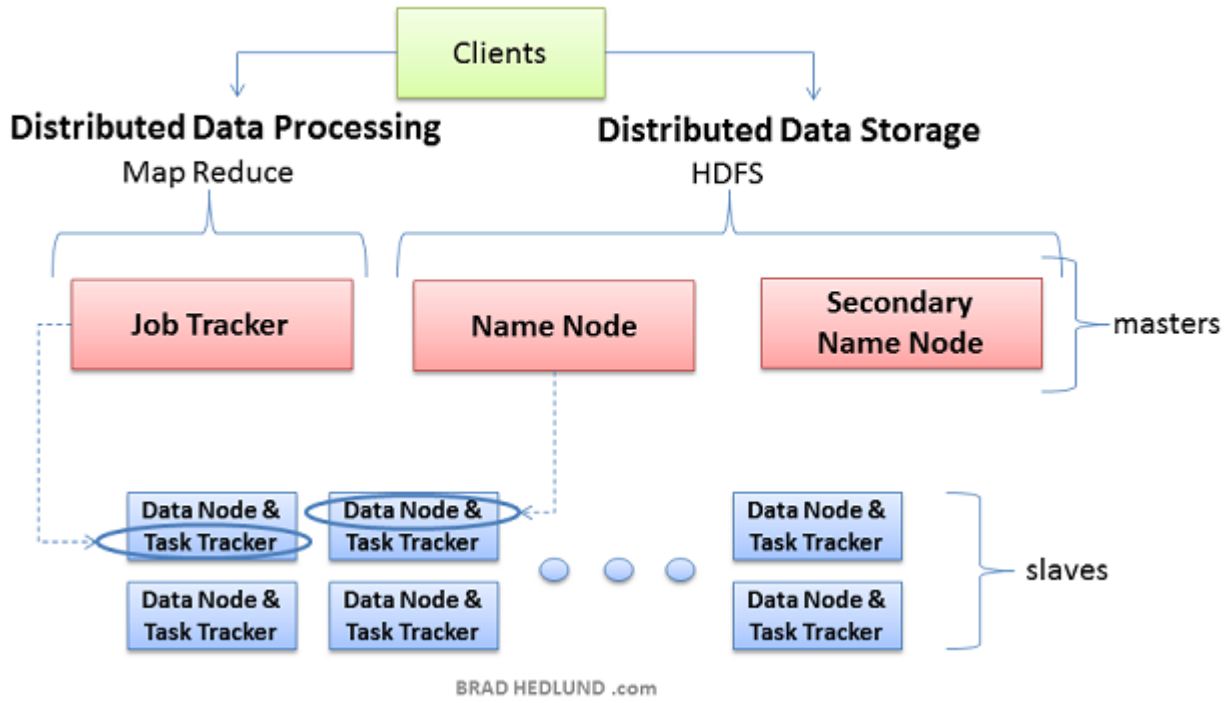
http://www.monitis.com/blog/2013/12/19/big-data-and-hadoop-whats-it-all-about/

Arquitectura



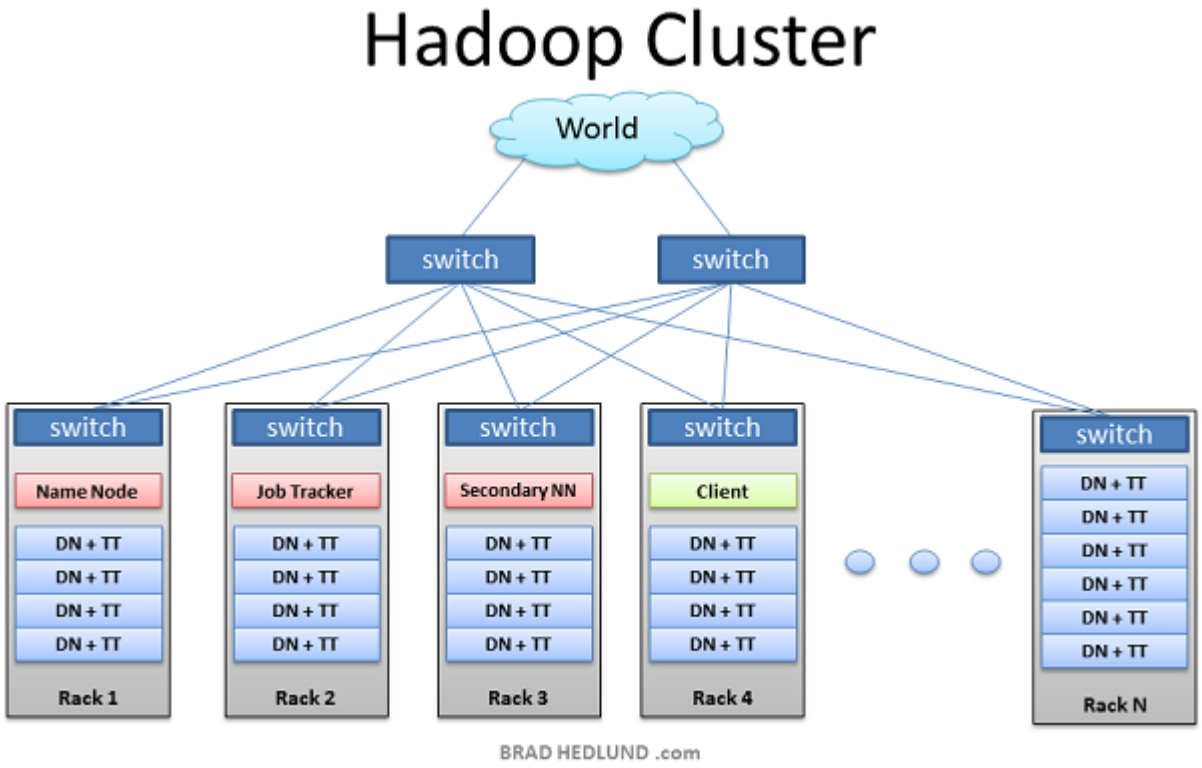
http://www.sachinpbuzz.com/2014/01/big-data-overview-of-apache-hadoop.html

Despliegue



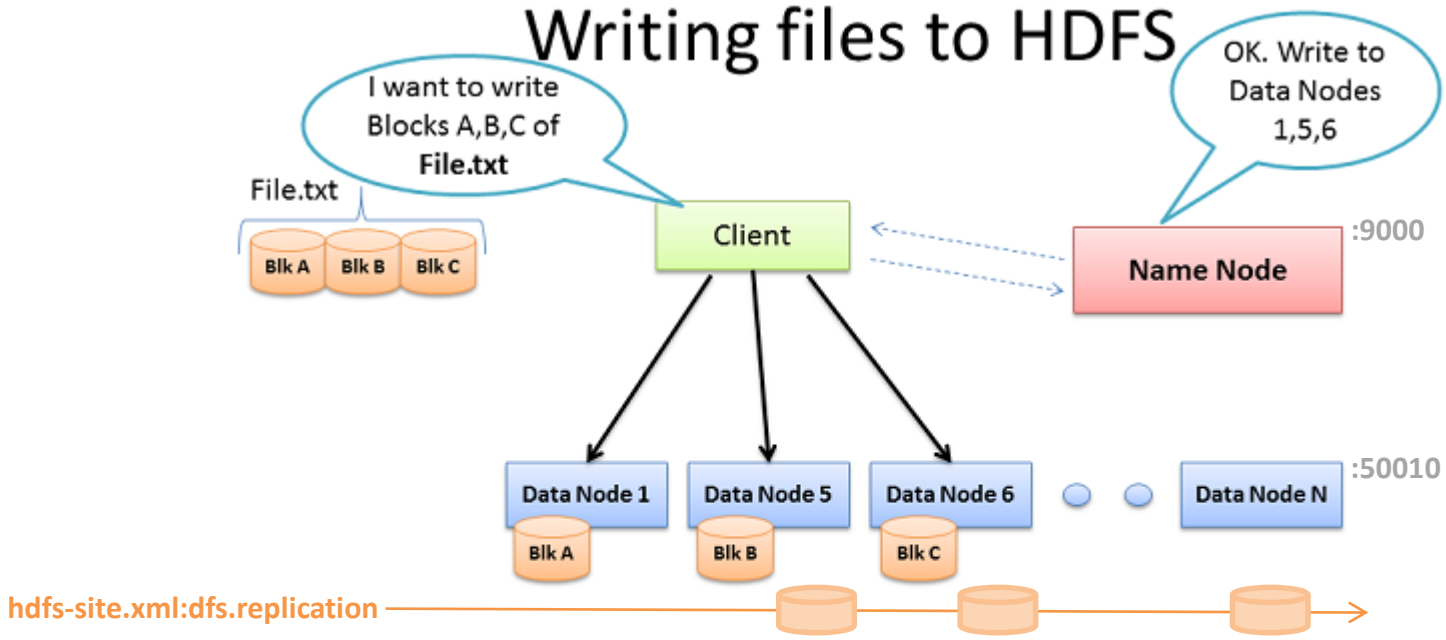
http://blog.csdn.net/suifeng3051/article/details/17288047

Despliegue



http://blog.csdn.net/suifeng3051/article/details/17288047

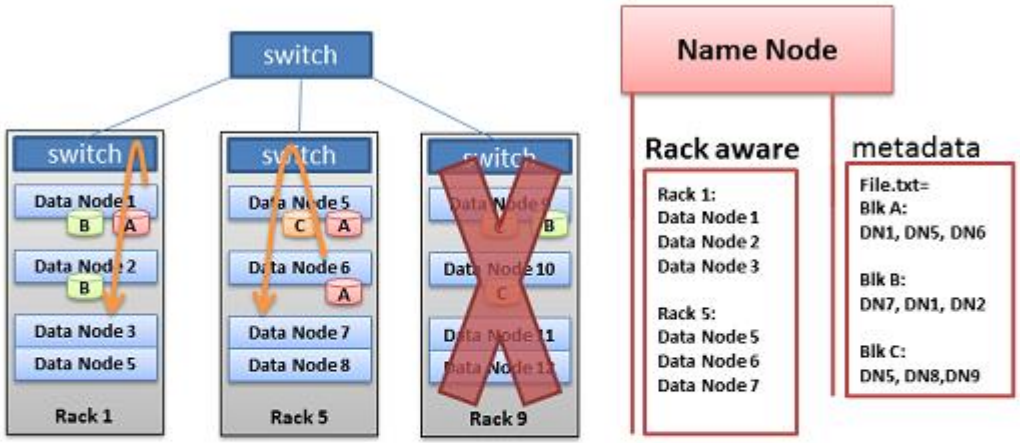
Despliegue



http://blog.csdn.net/suifeng3051/article/details/17288047

Despliegue

Hadoop Rack Awareness – Why?



- Never loose all data if entire rack fails
- Keep bulky flows in-rack when possible
- Assumption that in-rack is higher bandwidth, lower latency

BRAD HEDLUND .com

http://blog.csdn.net/suifeng3051/article/details/17288047

Despliegue

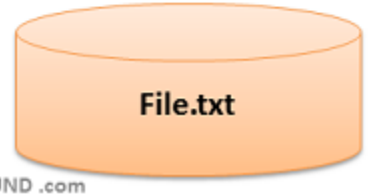
Typical Workflow

- Load data into the cluster (HDFS writes)
- Analyze the data (Map Reduce)
- Store results in the cluster (HDFS writes)
- Read the results from the cluster (HDFS reads)

Sample Scenario:

How many times did our customers type the word "**Refund**" into emails sent to customer service?

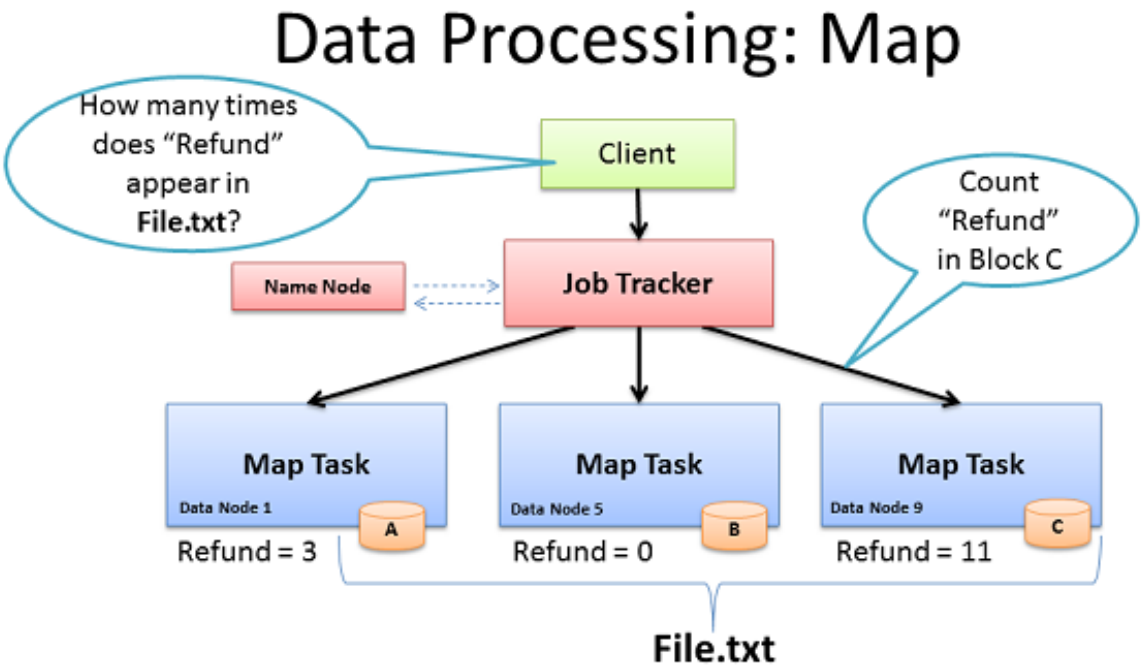
Huge file containing all emails sent to customer service



BRAD HEDLUND .com

http://blog.csdn.net/suifeng3051/article/details/17288047

Despliegue



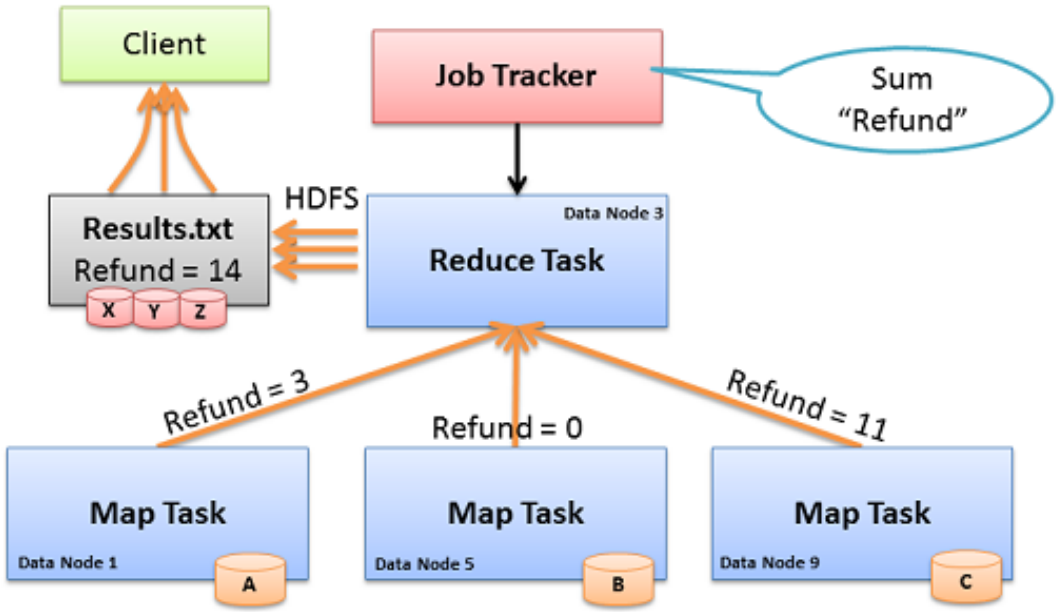
- **Map:** "Run this computation on your local data"
- Job Tracker delivers Java code to Nodes with local data

BRAD HEDLUND .com

http://blog.csdn.net/suifeng3051/article/details/17288047

Despliegue

Data Processing: Reduce



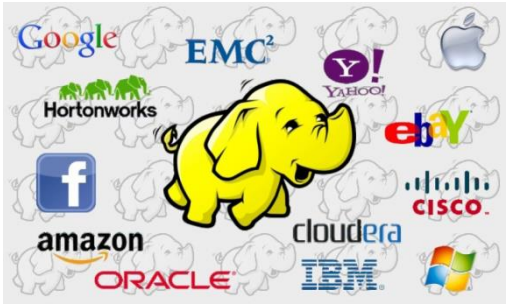
- **Reduce:** "Run this computation across Map results"
- Map Tasks send output data to Reducer over the network
- Reduce Task data output written to and read from HDFS

BRAD HEDLUND .com

http://blog.csdn.net/suifeng3051/article/details/17288047

Contenidos

Big
Big
Big
Big
Big

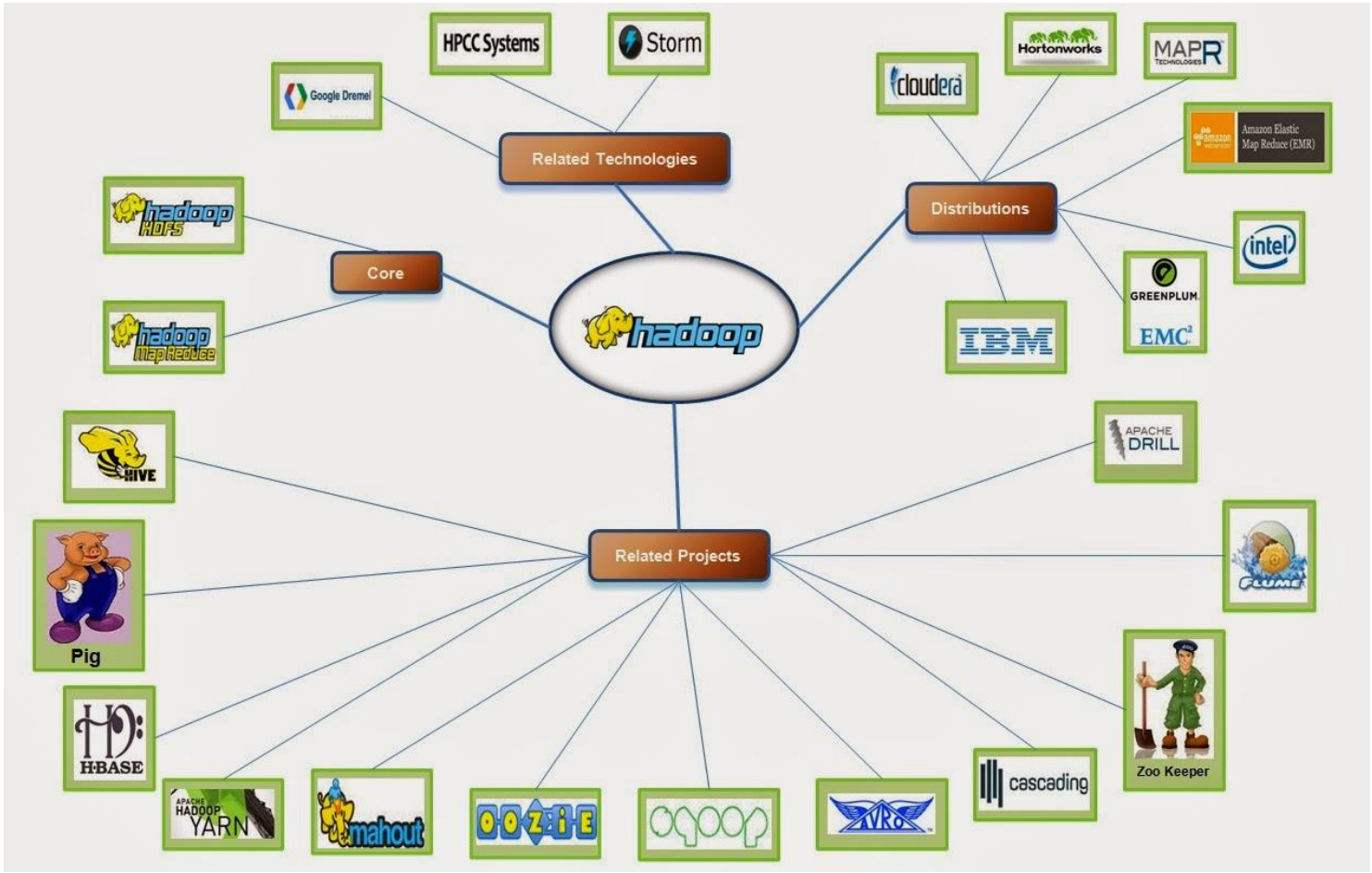


- Inicio
- Transición
- (re)Evolución
- (re)Retos

- Inicio
- Hadoop
- **Ecosistema**
- Evolución

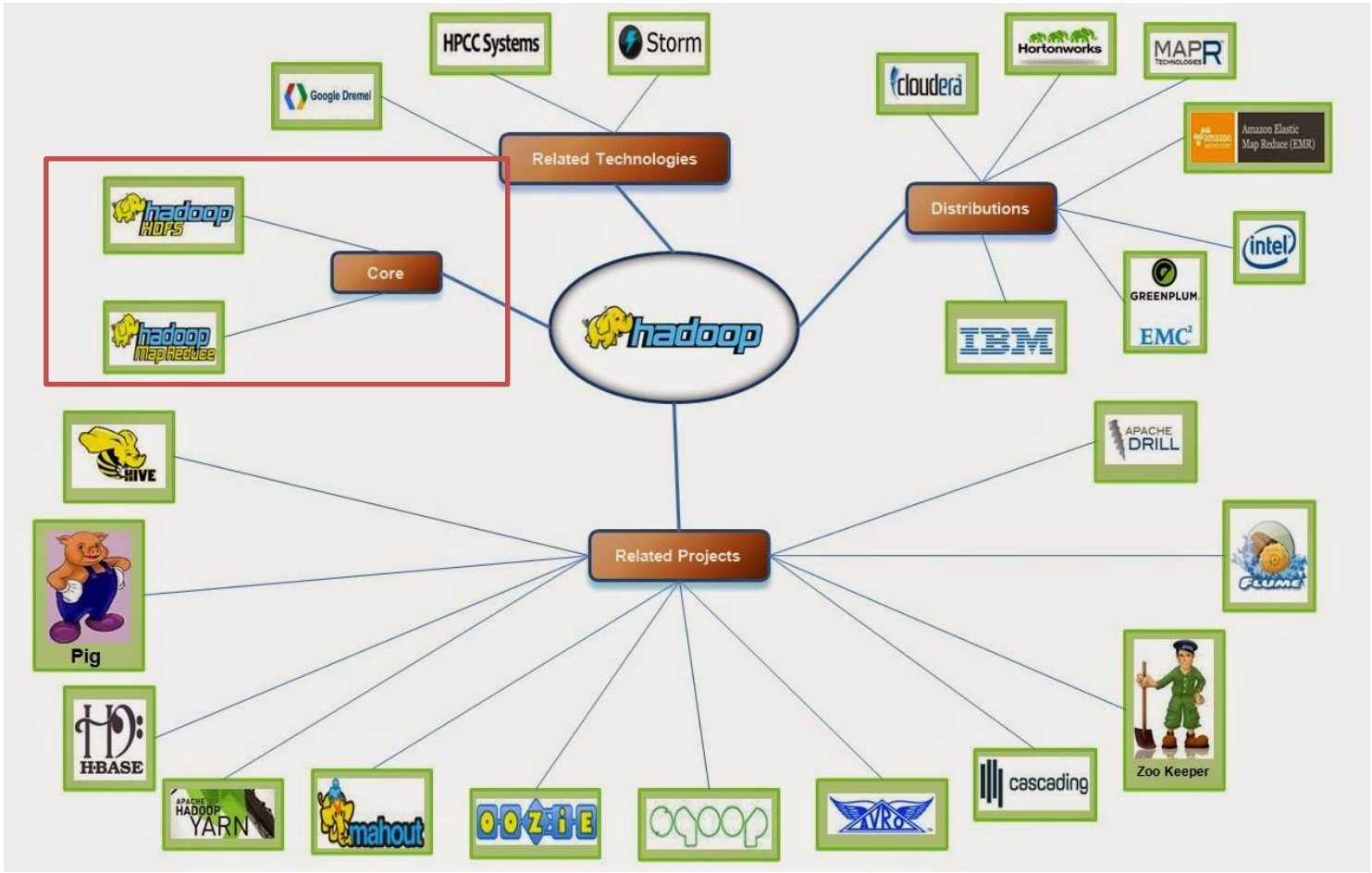
http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838
http://gruposda.es/wp-content/uploads/sertecni.jpg

Ecosistema



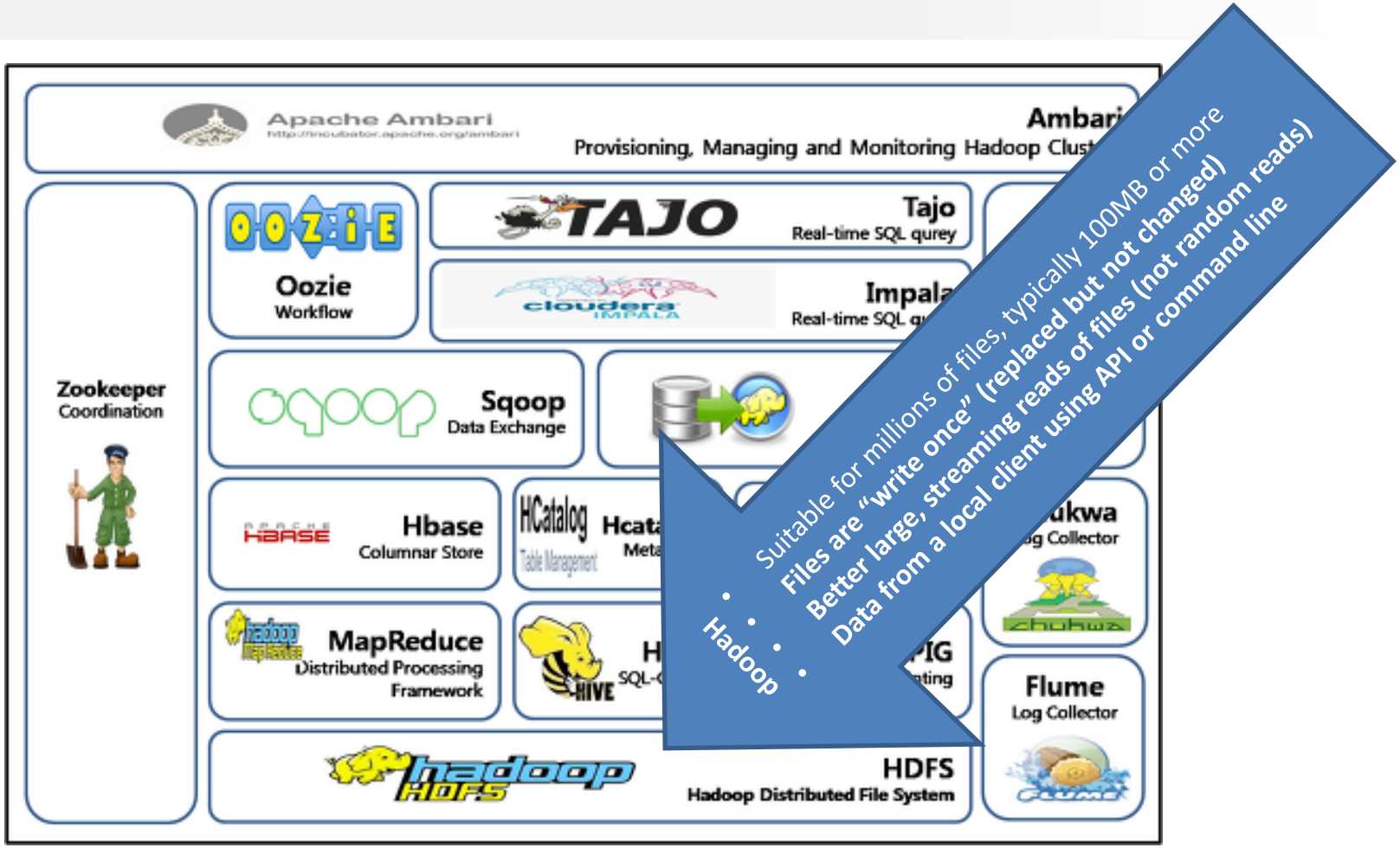
http://ambuj4bigdata.blogspot.com.es/2014_05_01_archive.html

Ecosistema



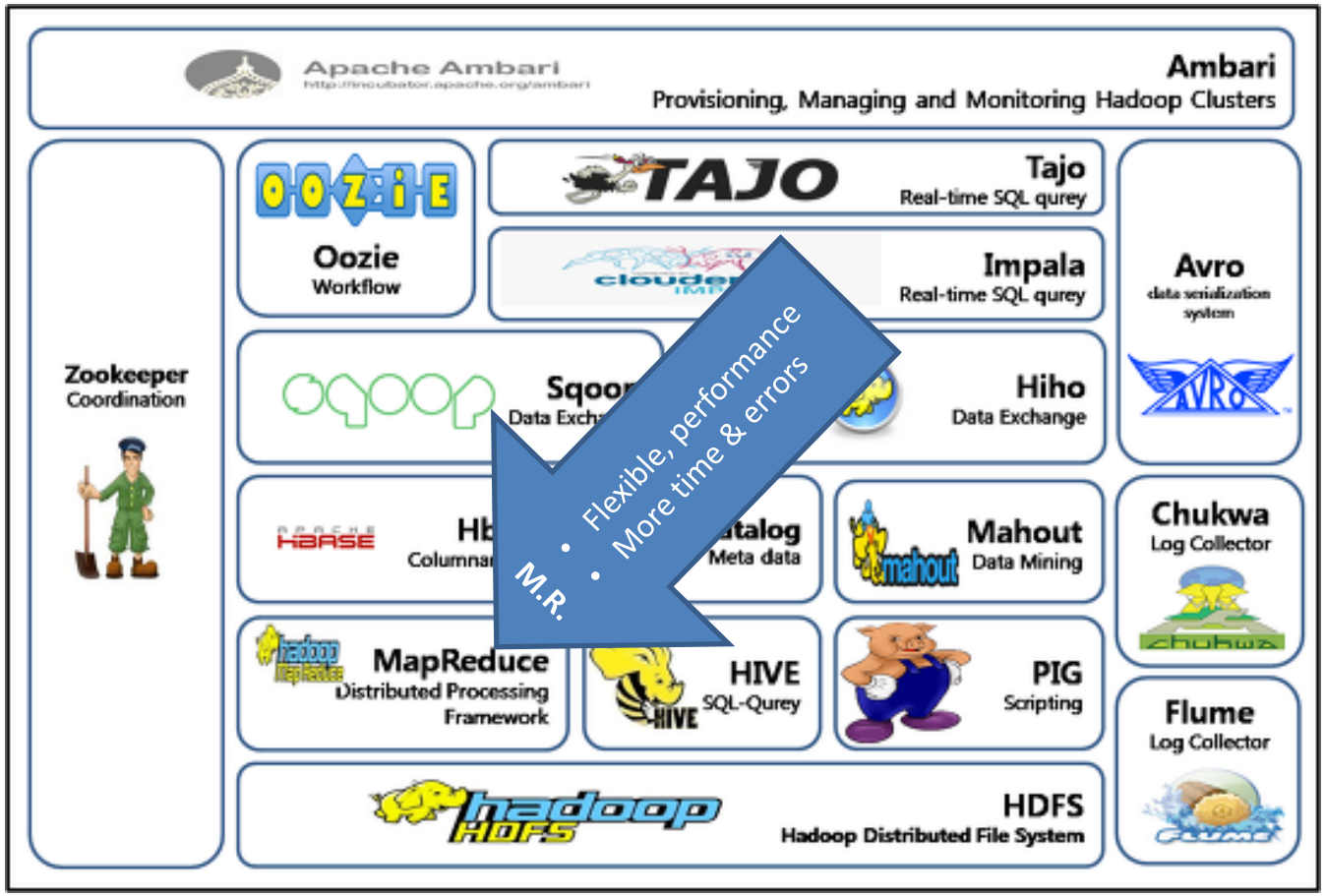
http://ambuj4bigdata.blogspot.com.es/2014_05_01_archive.html

Almacenamiento



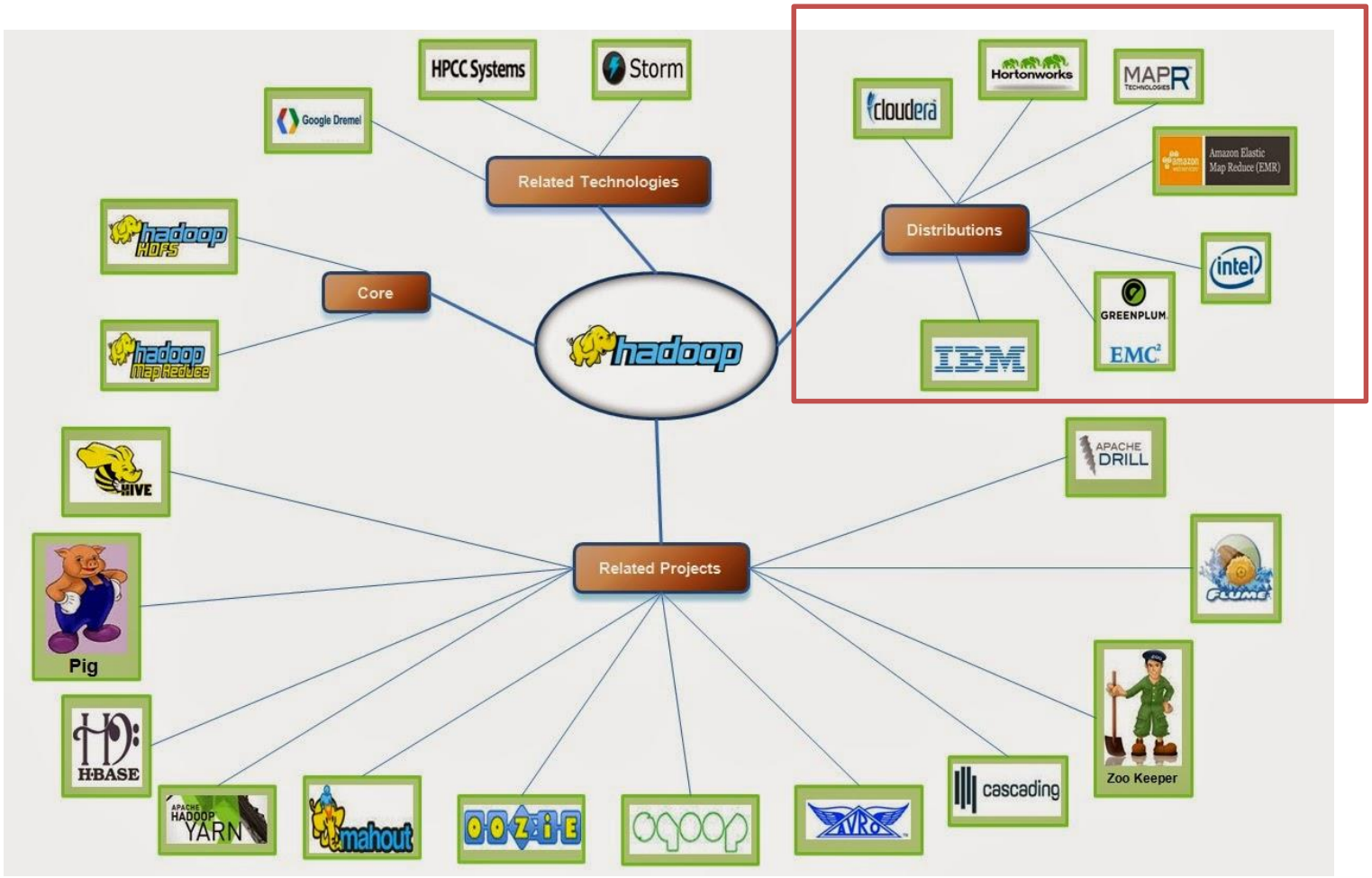
http://www.nextree.co.kr/p2865/

Procesamiento



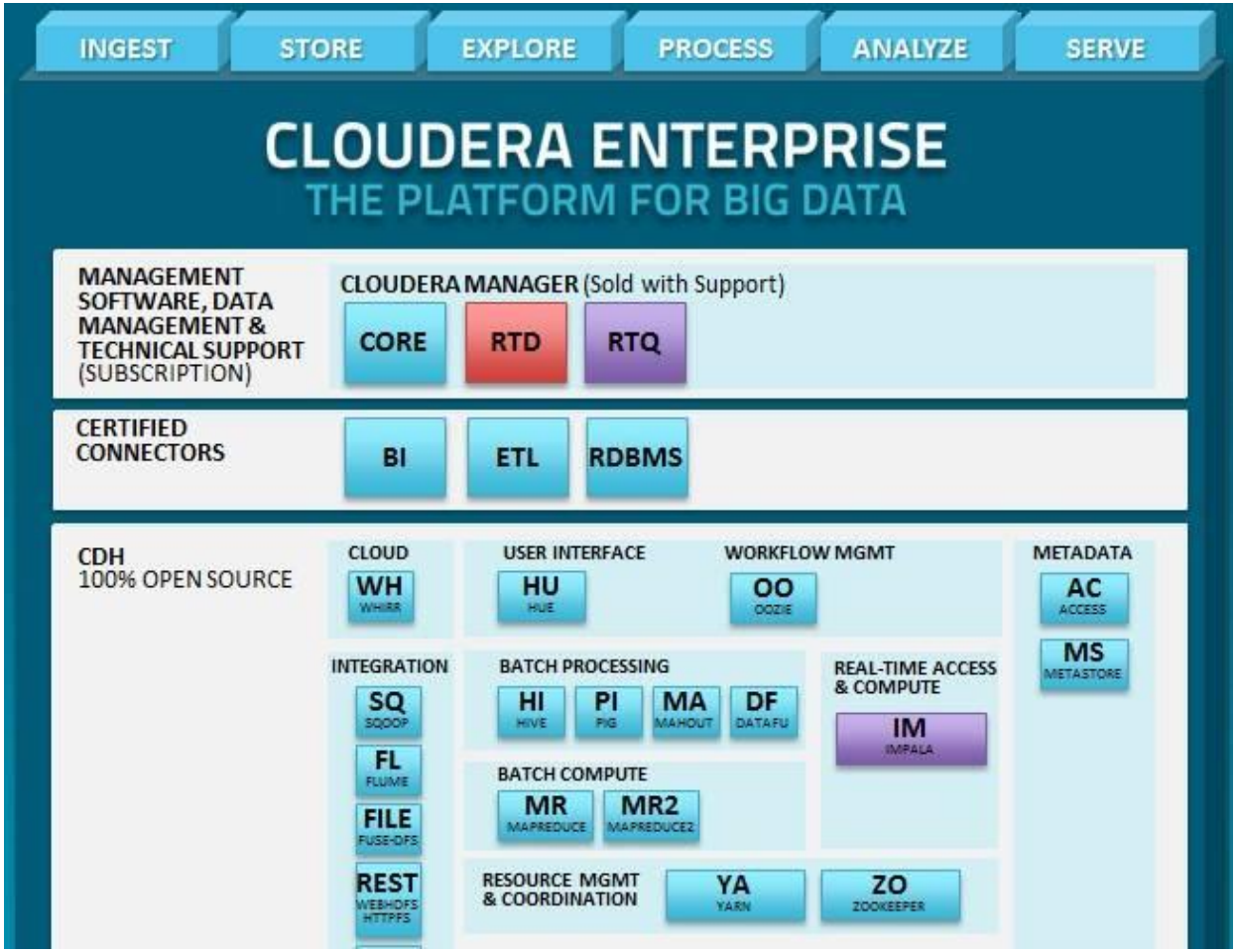
http://www.nextree.co.kr/p2865/

Ecosistema



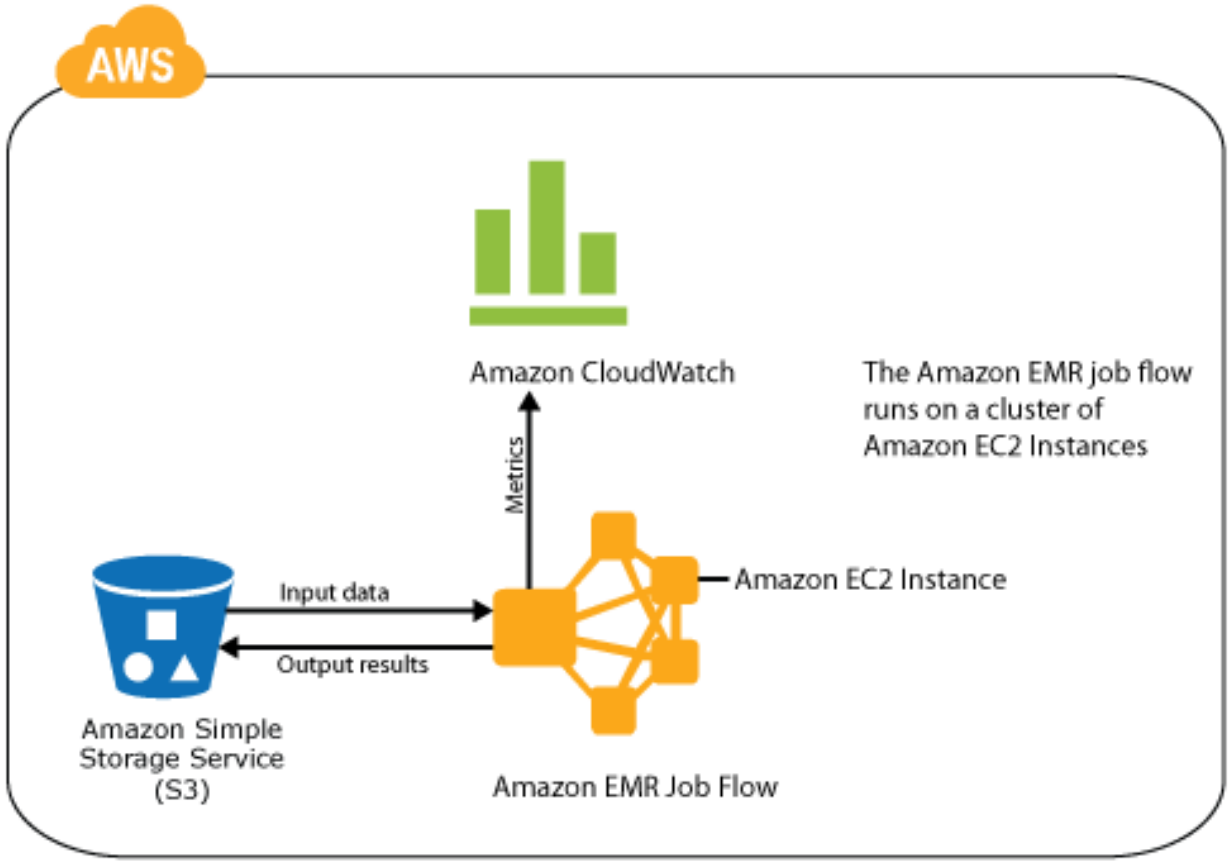
http://ambuj4bigdata.blogspot.com.es/2014_05_01_archive.html

CDH (Cloudera Hadoop Distribution)



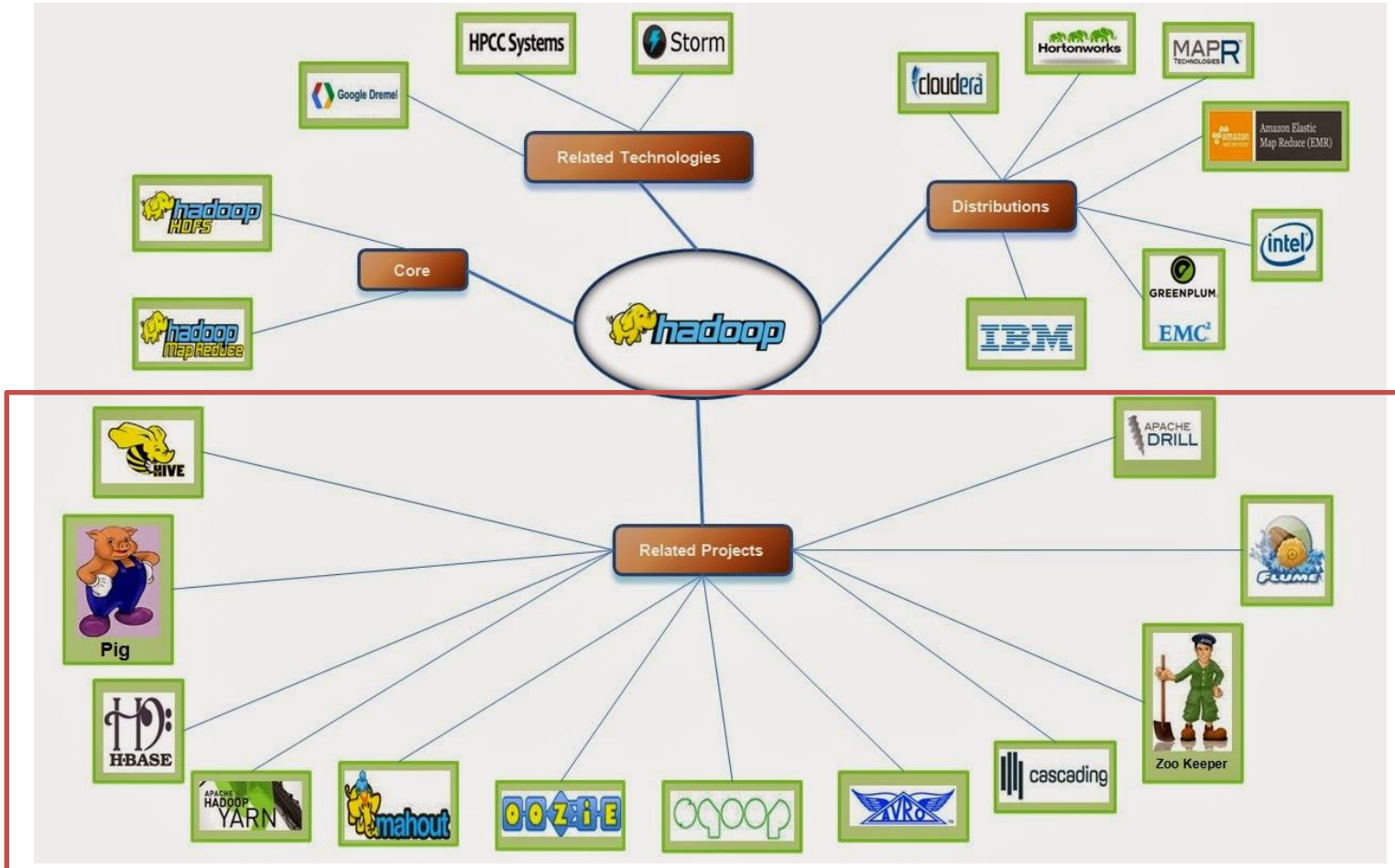
http://www.theregister.co.uk/2012/06/05/cloudera_cdh4_hadoop_stack/

AEMR (Amazon Elastic MapReduce)



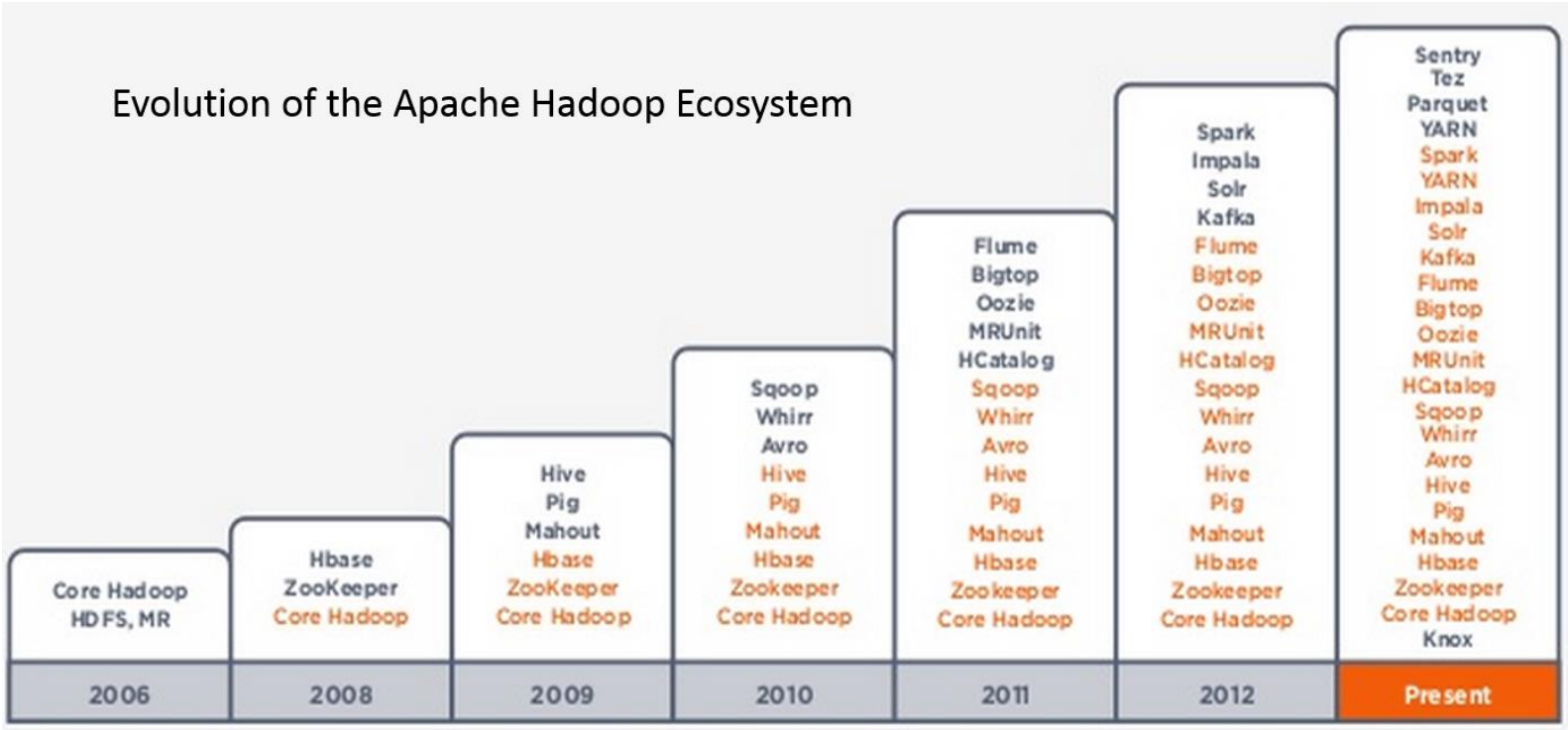
http://aws.amazon.com/es/elasticmapreduce/

Ecosistema



http://ambuj4bigdata.blogspot.com.es/2014_05_01_archive.html

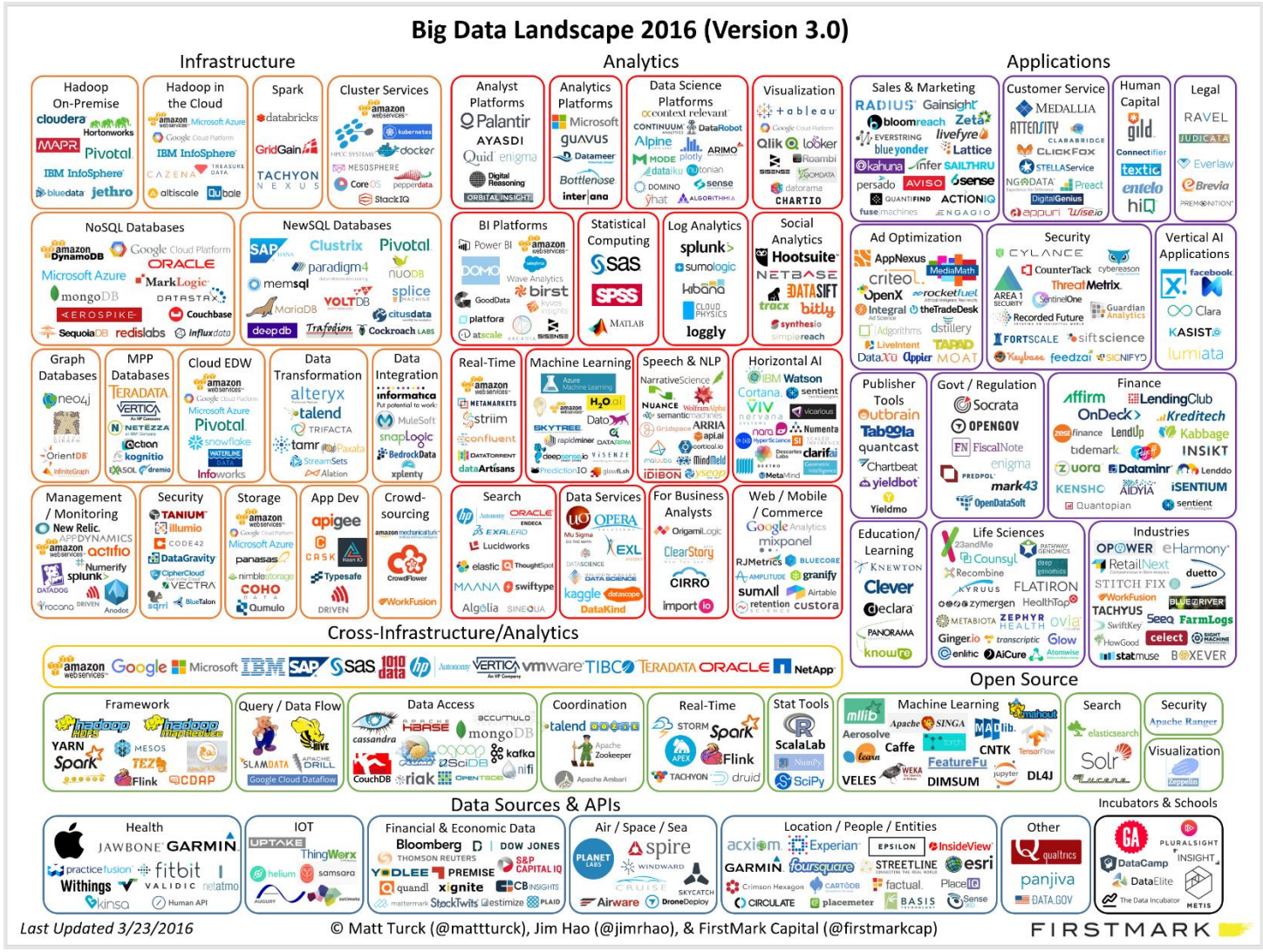
Ecosistema



https://practicalanalytics.files.wordpress.com/2011/11/hadoopevolution.png

Ecosistema

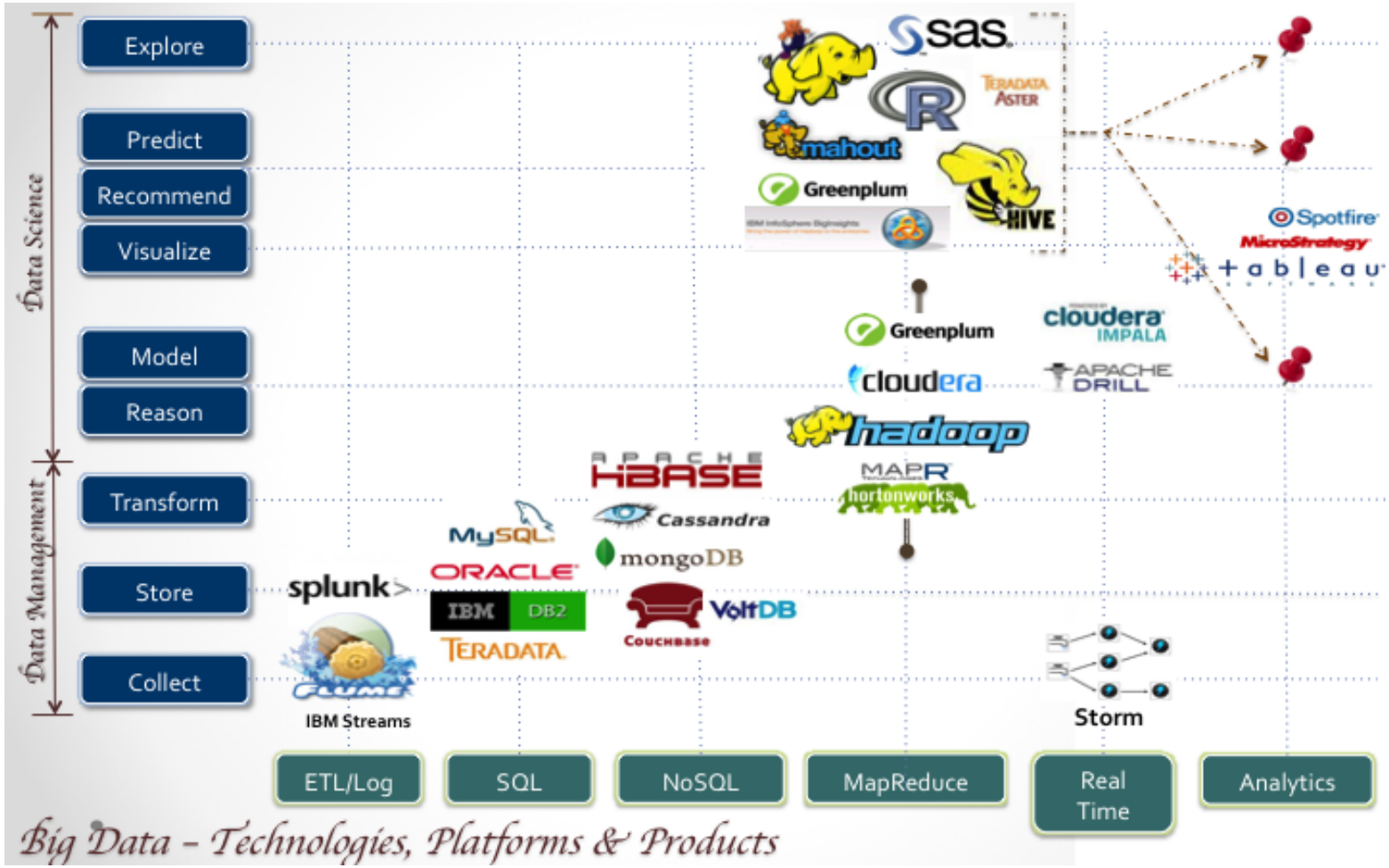
(funcionalidad)



http://mattturck.com/wp-content/uploads/2016/03/Big-Data-Landscape-2016-v18-FINAL.png + http://dfkoz.com/big-data-landscape/

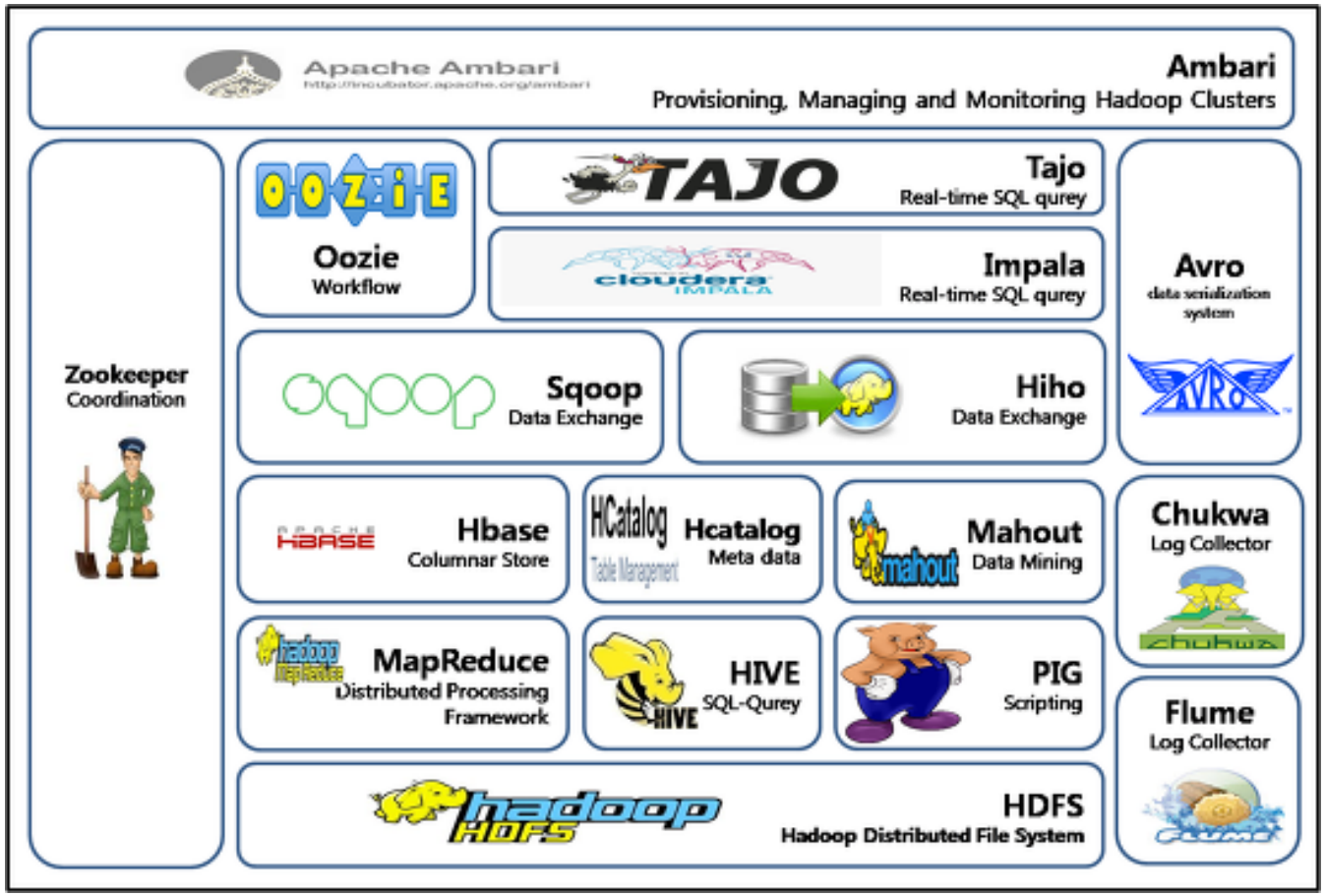
Ecosistema

(funcionalidad)



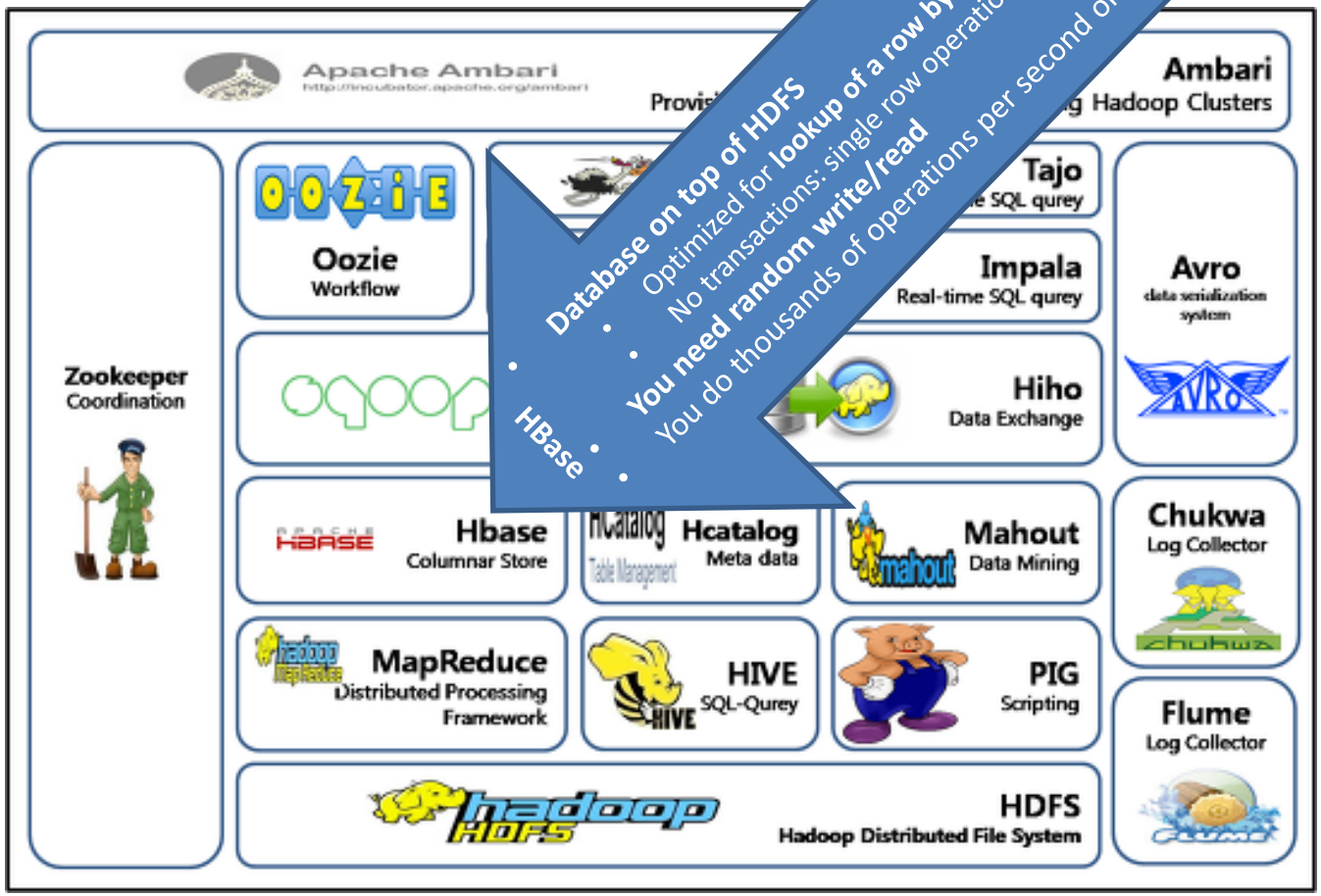
https://s-media-cache-ak0.pinimg.com/originals/0d/4f/0d/0d4f0d8aad9d144c52c696c603b8a27c.jpg

Ecosistema (estructura)



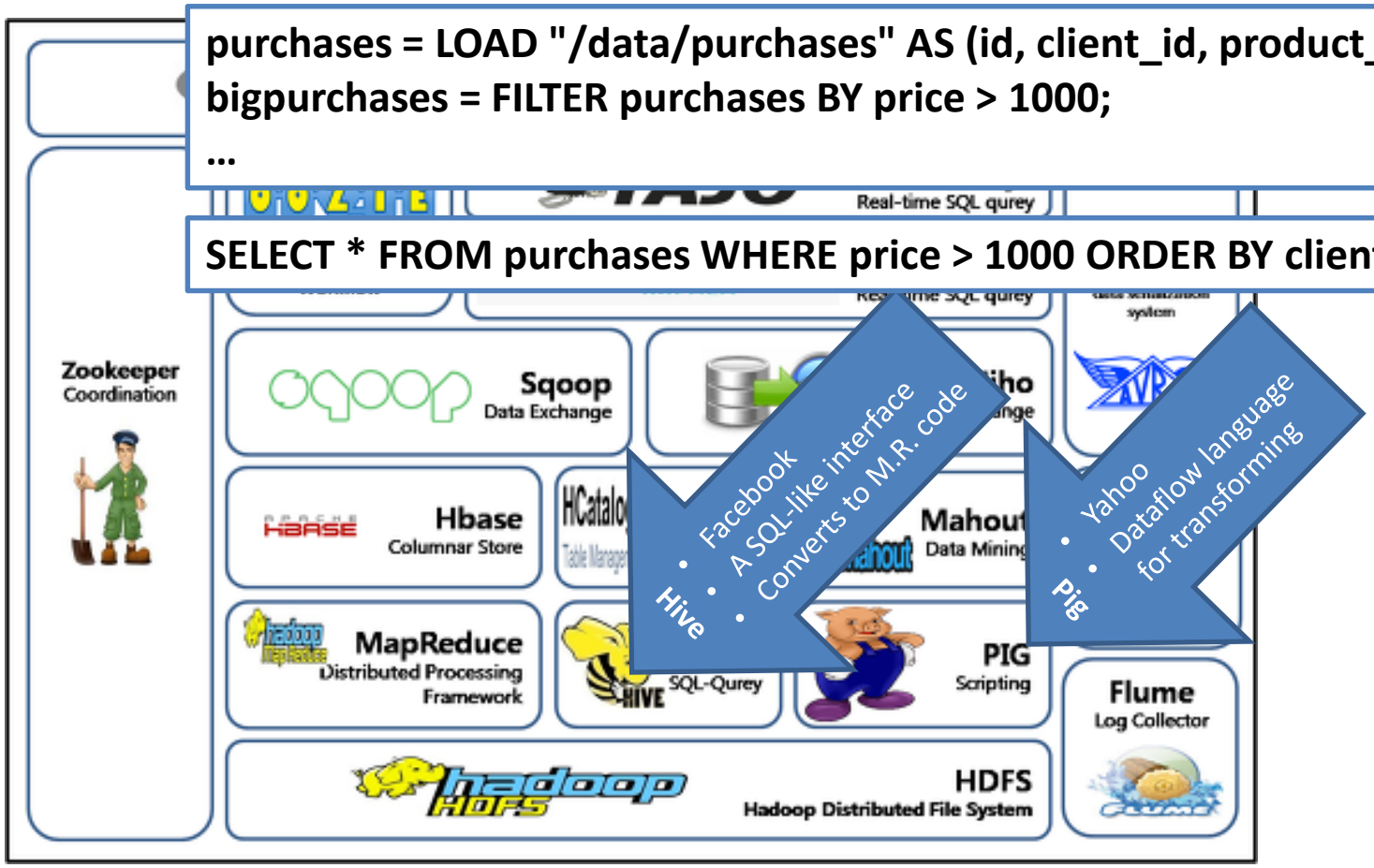
http://www.nextree.co.kr/p2865/

Almacenamiento



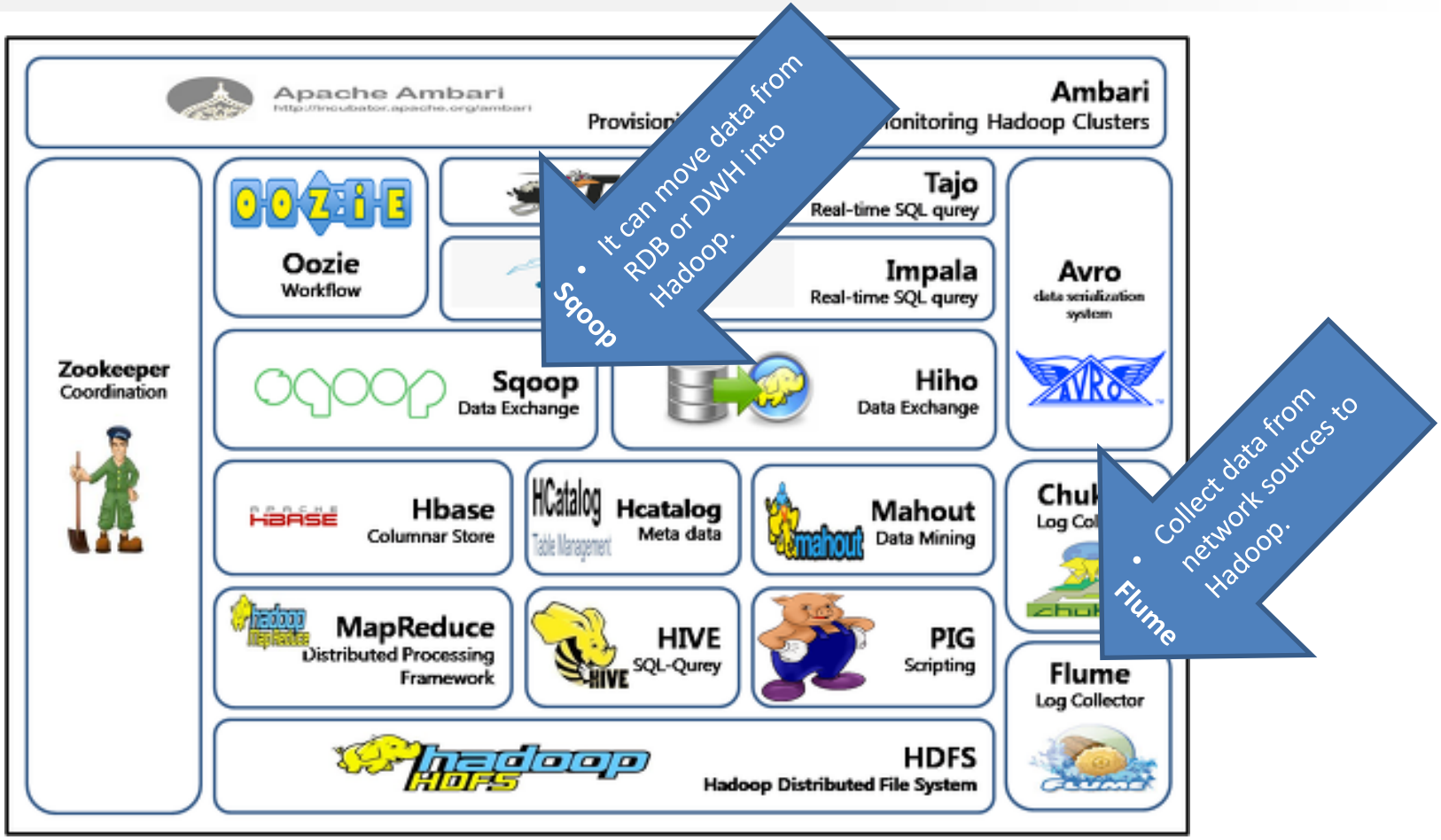
http://www.nextree.co.kr/p2865/

Procesamiento



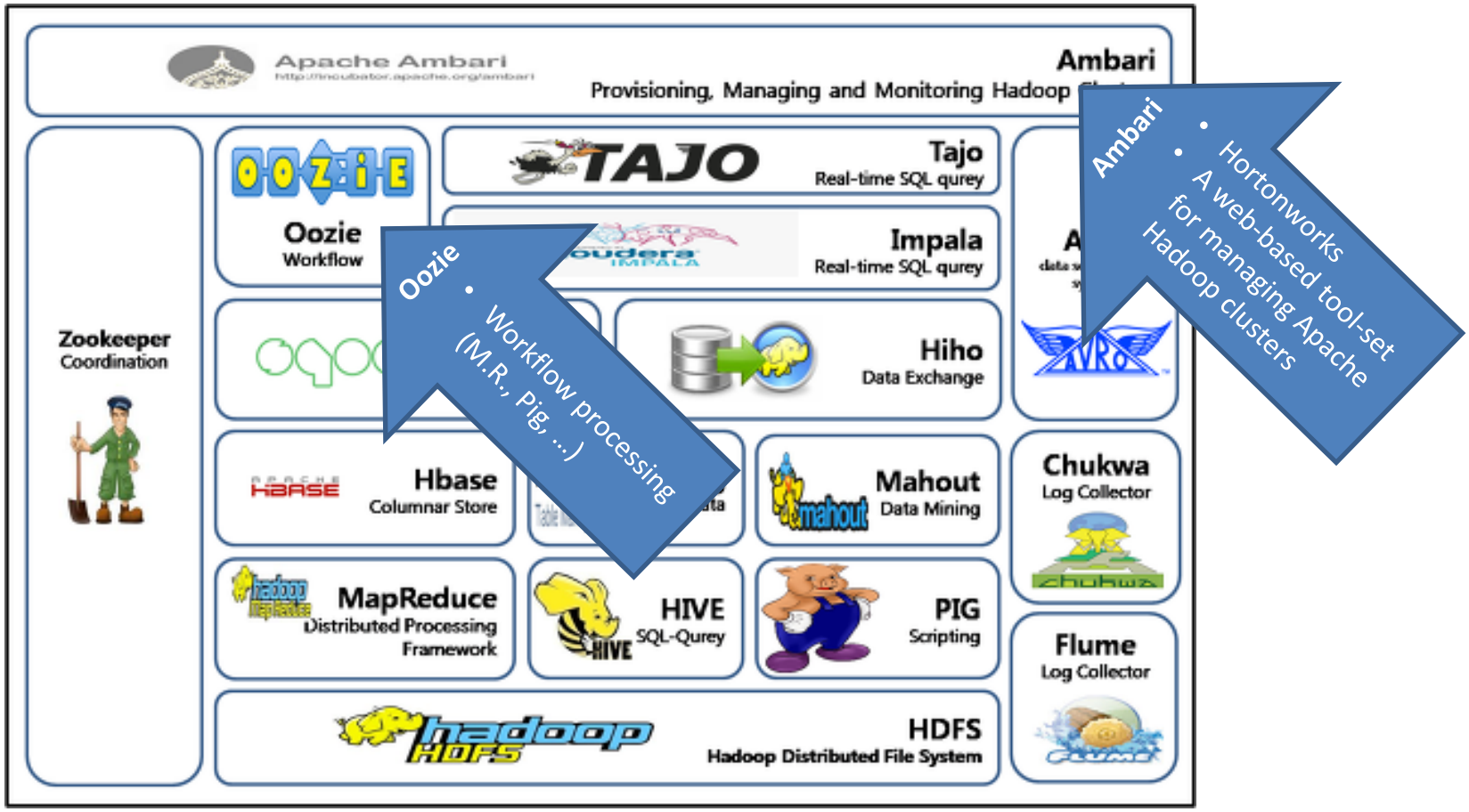
http://www.nextree.co.kr/p2865/

Almacenamiento: Integración



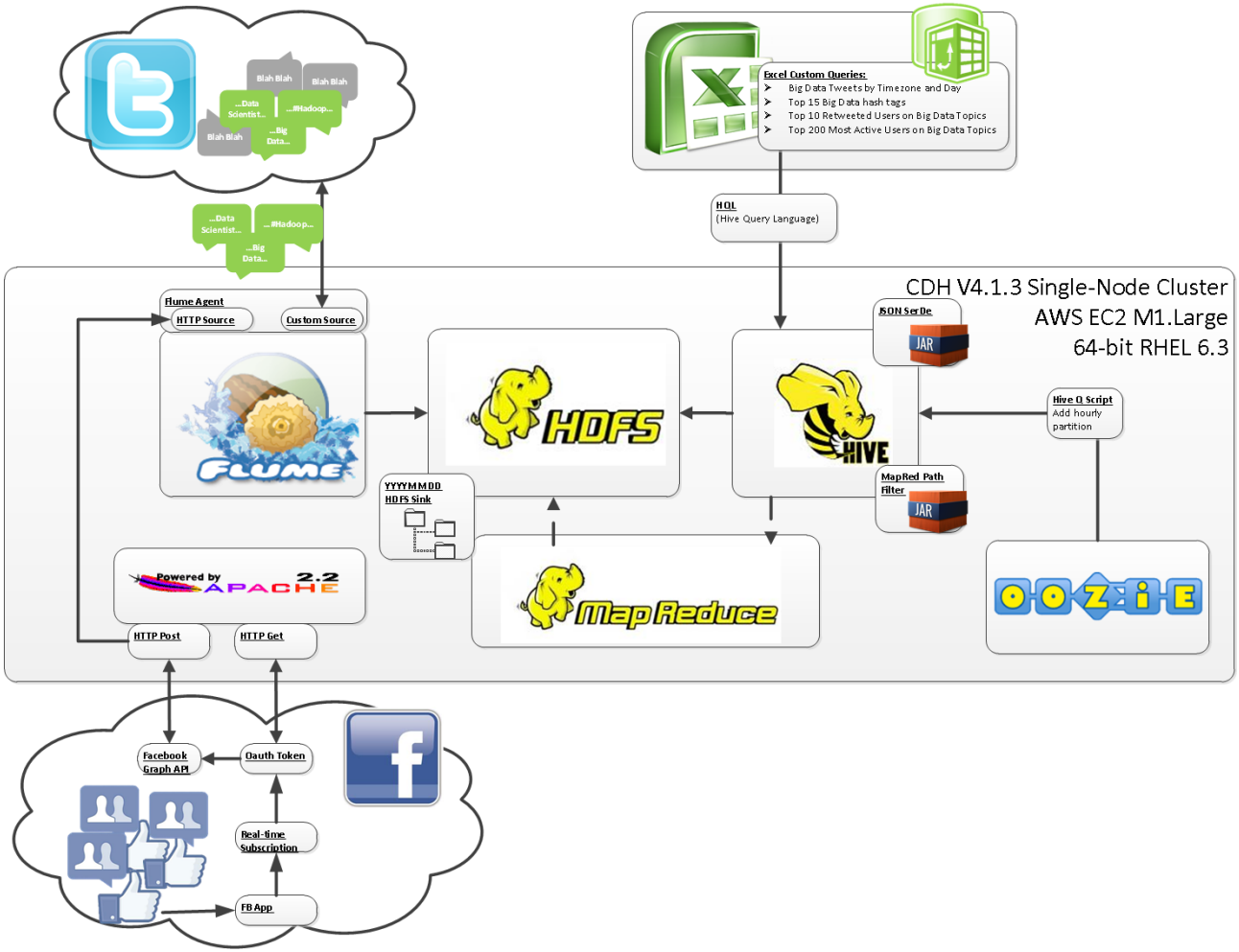
http://www.nextree.co.kr/p2865/

Procesamiento: orquestación



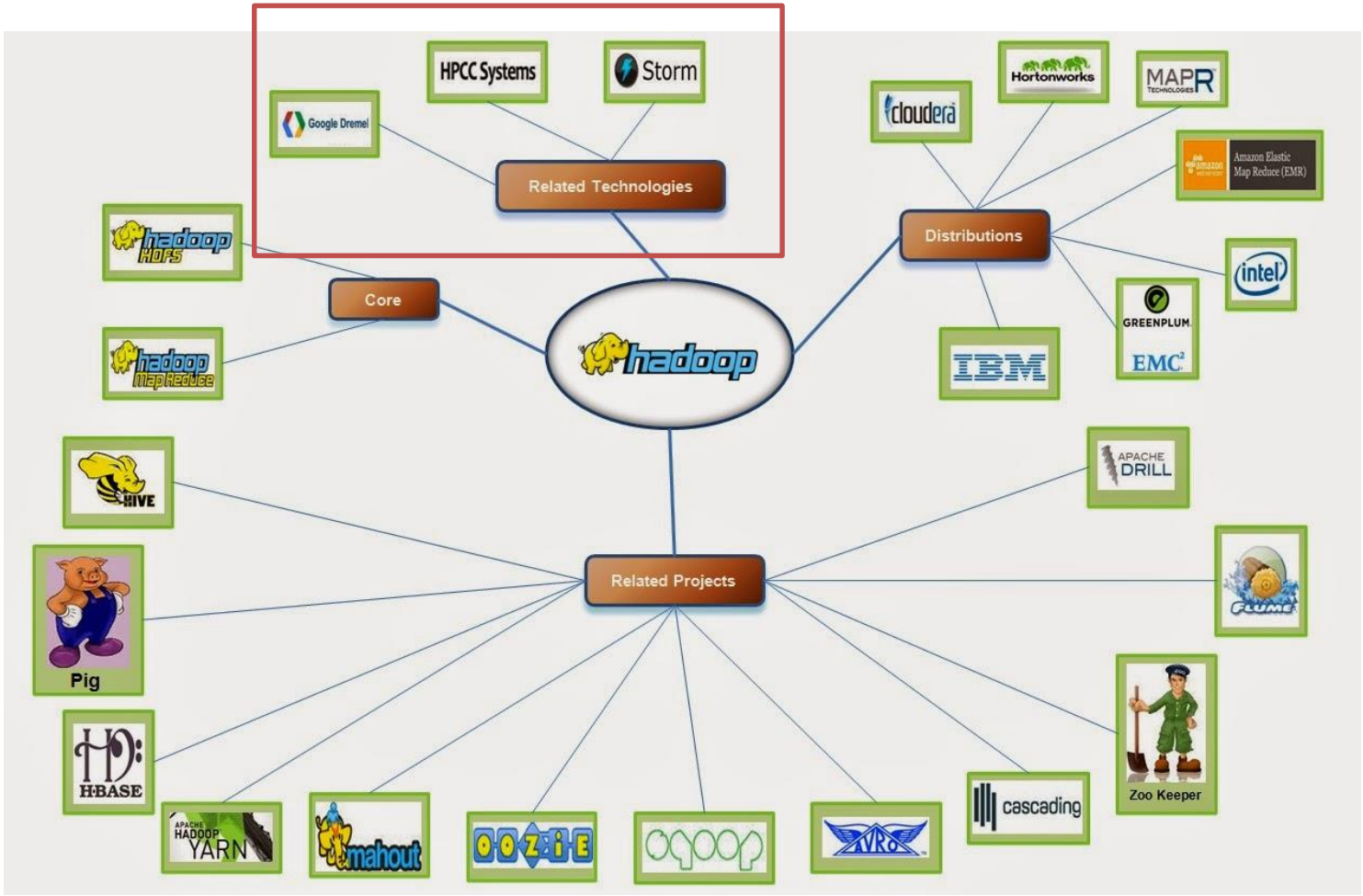
http://www.nextree.co.kr/p2865/

Ej.: Flume-Hive-oozie



http://www.datadansandler.com/2013/03/big-data-and-hadoop-assets-available-on.html

Ecosistema



http://ambuj4bigdata.blogspot.com.es/2014_05_01_archive.html

Contenidos

Big
Big
Big
Big
Big

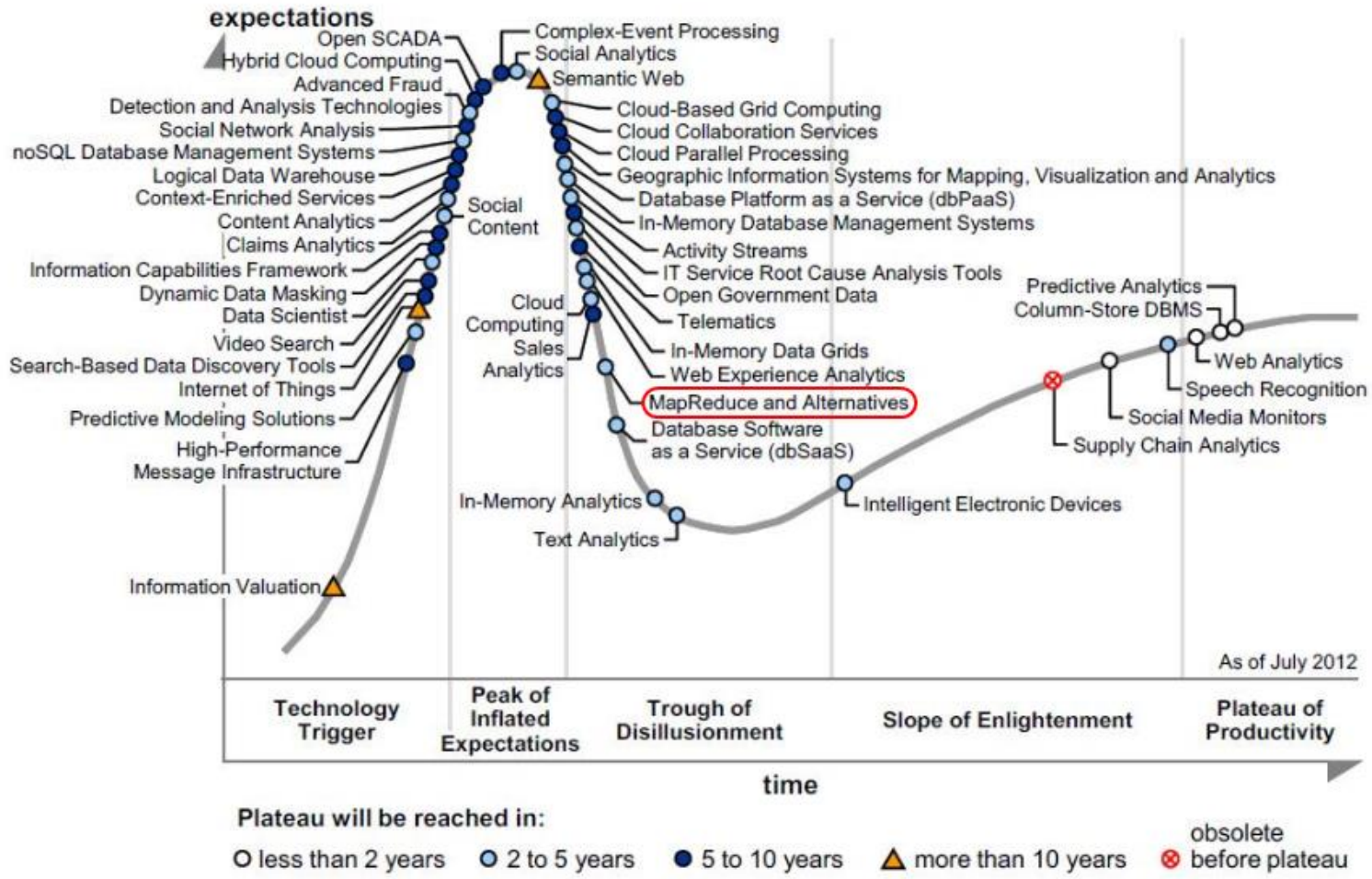
- Inicio
- Transición
- (re)Evolución
- (re)Retos

- Inicio
- Hadoop
- Ecosistema
- **Evolución**

http://www.siliconweek.es/wp-content/uploads/2013/08/BigData-datos-guardar-almacenamiento-fichero-archivo.jpg
http://datameer2.datameer.com/blog/wp-content/uploads/2012/06/Hadoop-Ecosystem-Infographic-21.png

Alternativas

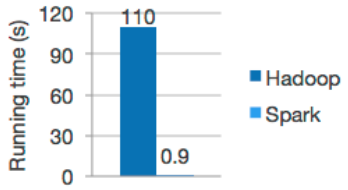
Figure 1. Hype Cycle for Big Data, 2012



Source: Gartner (July 2012)

https://www.simple-talk.com/cloud/data-science/analyze-big-data-with-apache-hadoop-on-windows-azure-preview-service-update-3/

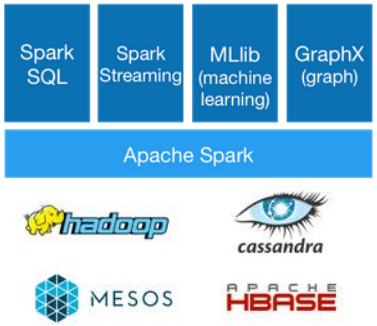
Spark



Logistic regression in Hadoop and Spark

```
file = spark.textFile("hdfs://...")

file.flatMap(lambda line: line.split())
    .map(lambda word: (word, 1))
    .reduceByKey(lambda a, b: a+b)
```

Word count in Spark's Python API



- Velocidad
  - Se evita en lo posible escribir en disco los resultados intermedios
- Facilidad de uso
  - Transformaciones y operaciones
- Polivalente
  - En Scala, Python, Java
  - Combina SQL, *streaming*, analítico…
  - Trabaja con HDFS, HBase, S3, …

https://spark.apache.org/

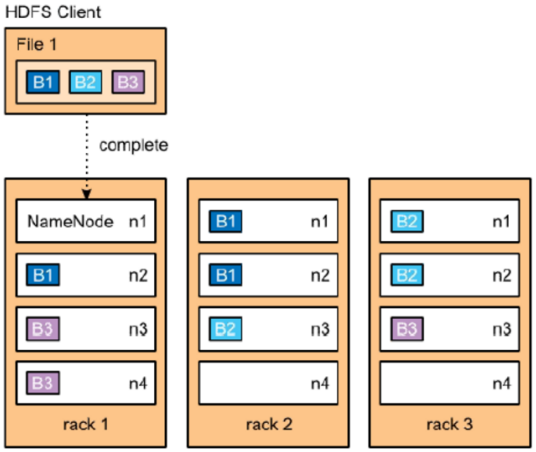
vs.

	Hadoop	Spark
Datos	Resultados intermedios en disco	Mantener en memoria todo lo posible
Tolerancia a fallos en datos	HDF (2+1)	**RDD**
Procesamiento	Trabajos map/reduce	**DAG**
Programación	Java + Otros con tuberías	Scala, Python, Java, etc.
Escenario	Trabajos lentos en *batch*	*Batch* + Real-time + Iterativo + Interactivo
	(Más) E/S de disco + red que Spark	(Más) memoria RAM que Hadoop
	Marco de trabajo de propósito general	Alternativa más que remplazamiento de Hadoop

http://aptuz.com/blog/is-apache-spark-going-to-replace-hadoop/

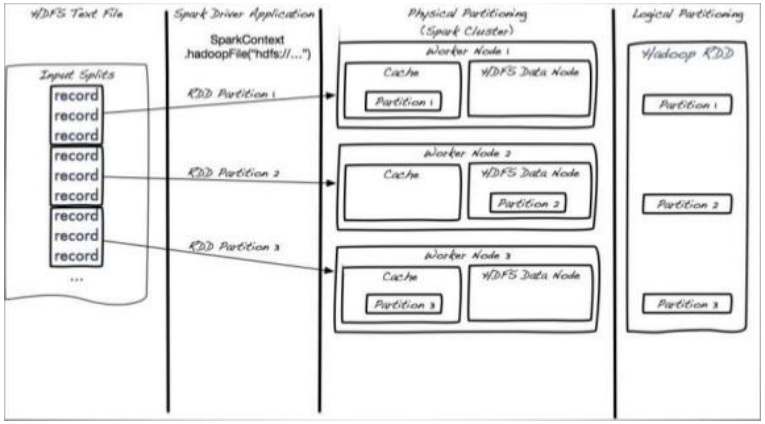
Spark vs. hadoop

HDFS
(HDFS + 3copies)



- Datos leídos en bloques
  - 64MB por defecto
- Usados como en ficheros

RDD
(*Resilient Distributed Dataset*)

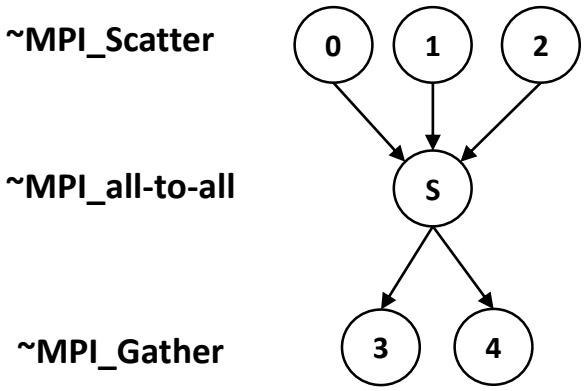


- Datos leídos como RDD
  - Inmutable, regenerable, en memoria
- DataFrame = RDD[fila] + Schema

https://geekytheory.com/apache-spark-que-es-y-como-funciona/
http://hadoopgeek.com/wp-content/uploads/2013/12/hdfswrite.png
http://www.slideshare.net/cfregly/spark-streaming-40659876

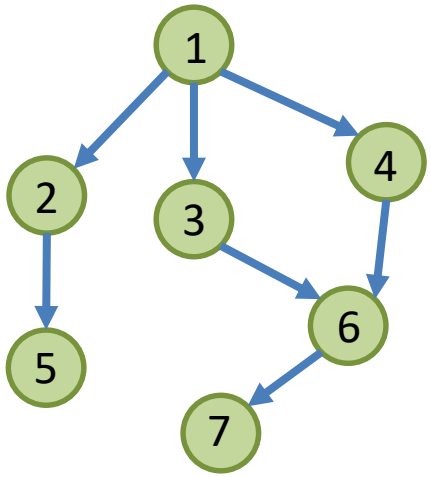
vs.

M-R
(Map-Reduce)



- DAG con 2 estados
  - Map + Reduce
- Datos intermedios a disco

DAG
(*Directed Acyclic Graph*)

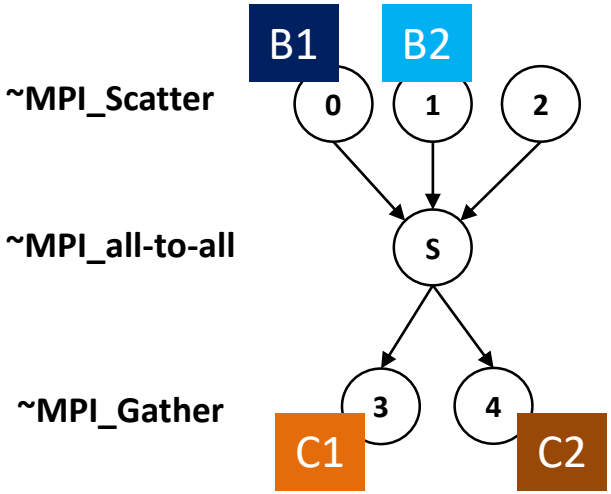


- DAG con N estados
- Datos intermedios en memoria

https://geekytheory.com/apache-spark-que-es-y-como-funciona/

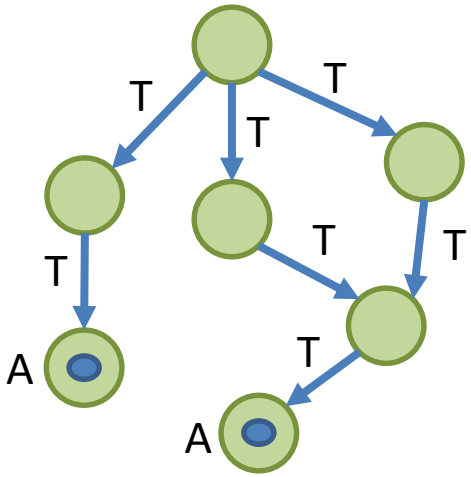
Spark vs. hadoop

HDFS + M-R



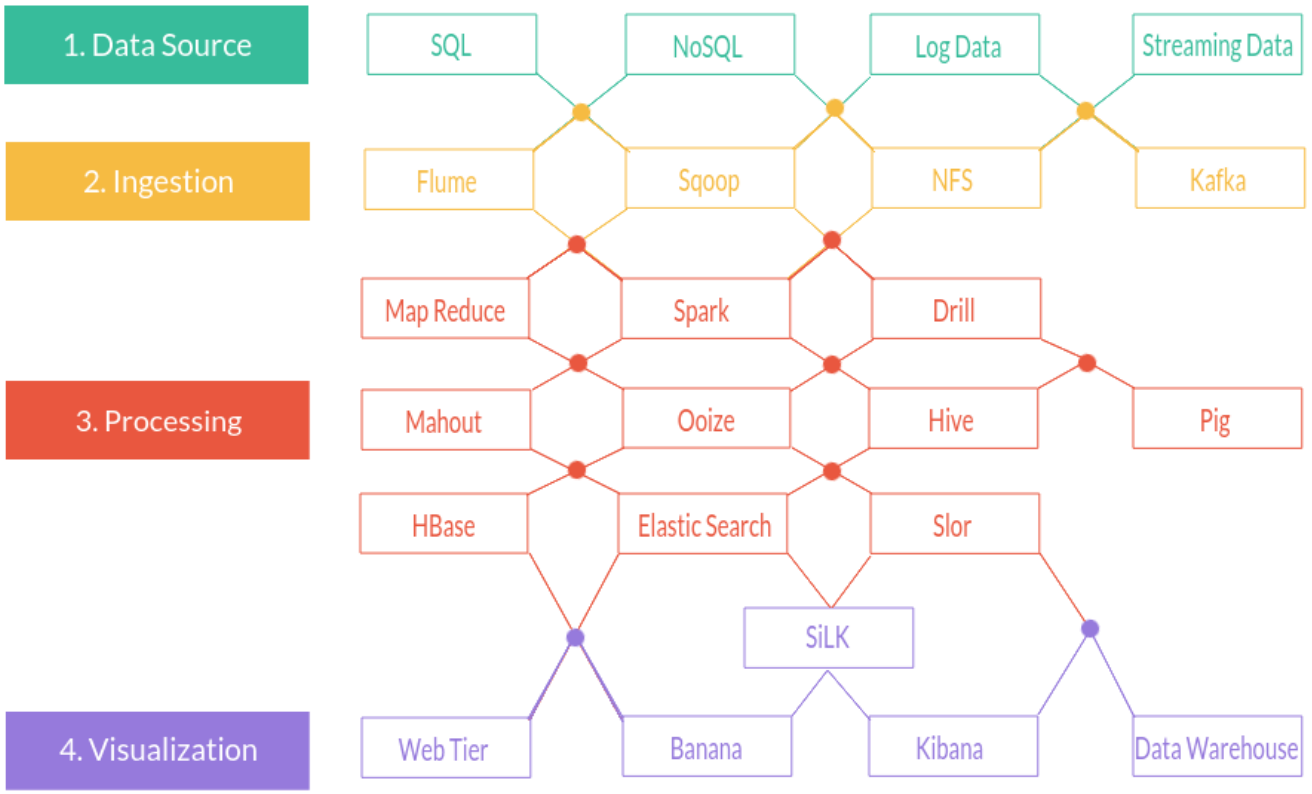
- Operaciones sobre bloques:
  - En M se leerán
  - En R se escribirán

RDD + DAG



- Dos tipos operaciones:
  - Transformaciones (RDD[] -> RDD'[])
  - Acciones (RDD[] -> valor)

https://geekytheory.com/apache-spark-que-es-y-como-funciona/



http://aptuz.com/blog/is-apache-spark-going-to-replace-hadoop/


```
1. val conf = new SparkConf().setMaster("local[2]")
2. val sc = new SparkContext(conf)
3. val lines = sc.textFile(path, 2)
4. val words = lines.flatMap(_.split(" "))
5. val pairs = words.map(word => (word, 1))
6. val wordCounts = pairs.reduceByKey(_ + _)
7. val localValues = wordCounts.take(100)
8. localValues.foreach(r => println(r))
```

https://www.youtube.com/watch?v=x8xXXqvhZq8

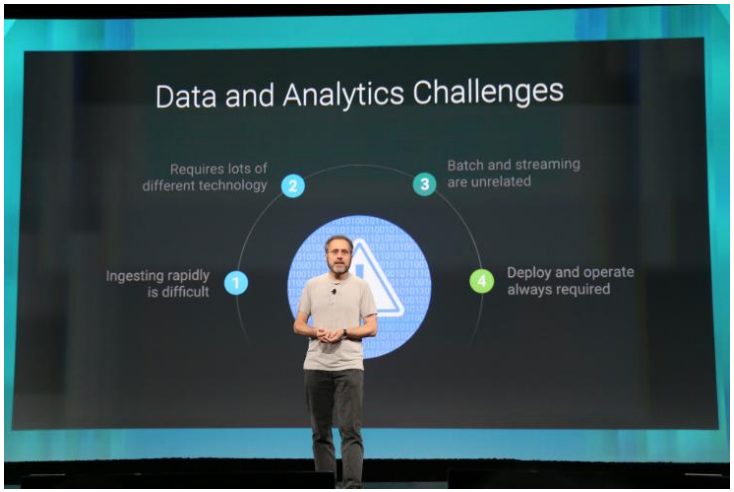

```
1. val conf = new SparkConf().setMaster("local[2]")
2. val sc = new SparkContext(conf)
3. val lines = sc.textFile(path, 2)
4. val words = lines.flatMap(_.split(" "))
5. val pairs = words.map(word => (word, 1))
6. val wordCounts = pairs.reduceByKey(_ + _)
7. val localValues = wordCounts.take(100)
8. localValues.foreach(r => println(r))
```

https://www.youtube.com/watch?v=x8xXXqvhZq8


```
1. val conf = new SparkConf().setMaster("local[2]")
2. val sc = new SparkContext(conf)
3. val lines = sc.textFile(path, 2)
4. val words = lines.flatMap(_.split(" "))
5. val pairs = words.map(word => (word, 1))
6. val wordCounts = pairs.reduceByKey(_ + _)
7. val localValues = wordCounts.take(100)
8. localValues.foreach(r => println(r))
```

https://www.youtube.com/watch?v=x8xXXqvhZq8

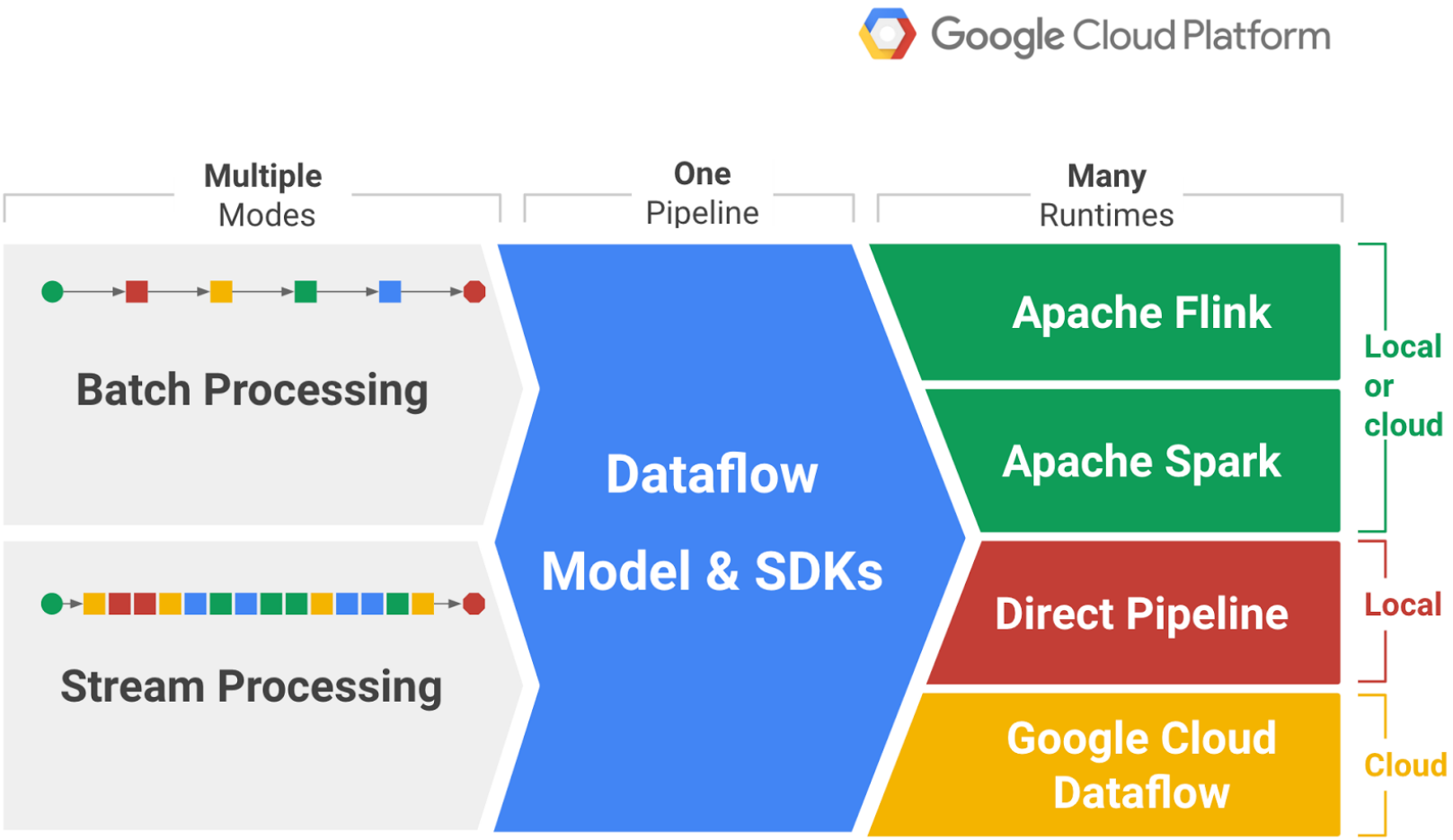
Google Dataflow



- Data pipelines for helping to ingest, transform and, analyze data.

http://techcrunch.com/2014/06/25/google-launches-cloud-dataflow-a-managed-data-processing-service/

Google Dataflow



https://4.bp.blogspot.com/-RlLeDymI_mU/Vp-1cb3AxNI/AAAAAAAACSQ/5TphliHJA4w/s1600/dataflow%2BASF.png

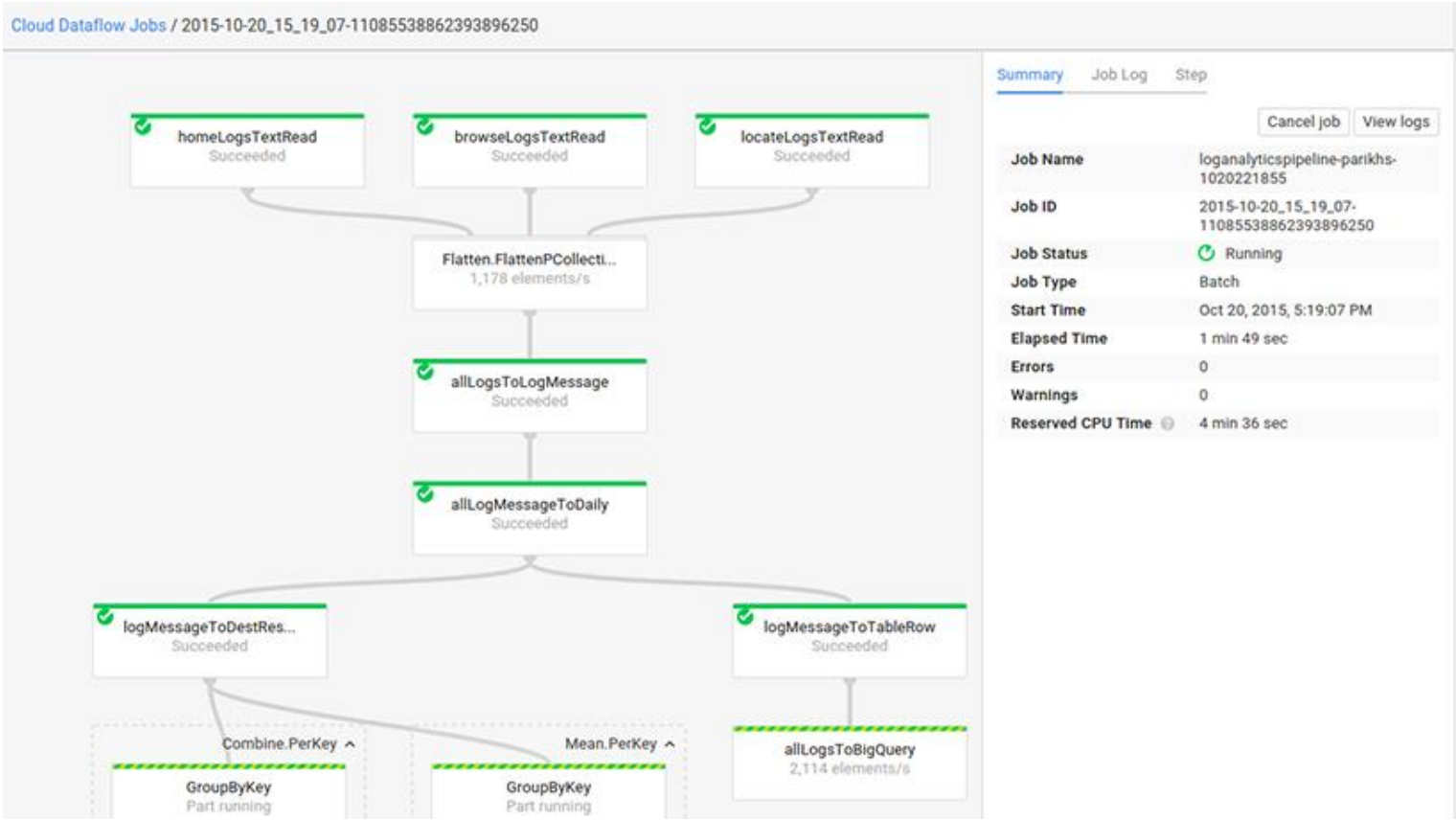
Google Dataflow


```
Pipeline p = Pipeline.create();
p.begin()

    .apply(TextIO.Read.from("gs://…"))

    .apply(ParDo.of(new ExtractTags()))

    .apply(Count.create())

    .apply(ParDo.of(new ExpandPrefixes())

    .apply(Top.larges

    .apply(TextIO.Wri

p.run();
```

```
class ExpandPrefixes … {
  ...
  public void processElement(ProcessContext c) {
    String word = c.element().getKey();
    for (int i = 1; i <= word.length(); i++) {
      String prefix = word.substring(0, i);
      c.output(KV.of(prefix, c.element()));
    }
  }
}
```

http://www.slideshare.net/GoogleCloudPlatformJP/google-cloud-dataflow-bqsushi

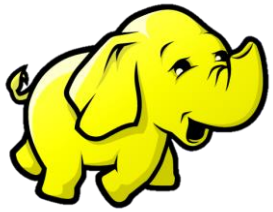
Google Dataflow



https://cloud.google.com/solutions/processing-logs-at-scale-using-dataflow

Contenidos

Big
Big
Big
Big
Big



- Inicio
- Transición
- (re)Evolución
- (re)Retos

- Inicio
- Hadoop
- Ecosistema
- Evolución

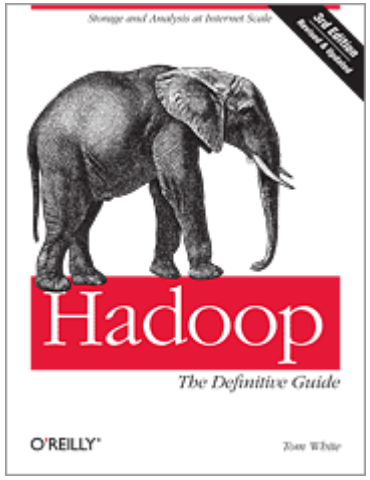
http://image.slidesharecdn.com/baronbigdatadeveloper-120418090010-phpapp02/95/the-big-data-developer-pavlobaron-1-728.jpg?cb=1338966838
http://gruposda.es/wp-content/uploads/sertecni.jpg

Bibliografía: tutoriales

- Página Web oficial:
  - http://hadoop.apache.org/
- Introducción a cómo funciona Hadoop:
  - http://blog.csdn.net/suifeng3051/article/details/17288047
- Tutorial de cómo instalar y usar Hadoop:
  - http://www.bogotobogo.com/Hadoop/BigData_hadoop_Install_on_ubuntu_single_node_cluster.php
  - http://www.bogotobogo.com/Hadoop/BigData_hadoop_Running_MapReduce_Job.php

Bibliografía: libro

- Hadoop: The Definitive Guide, 3rd Edition:
  - http://shop.oreilly.com/product/0636920021773.do
  - https://github.com/tomwhite/hadoop-book/



Bibliografía: TFG

- Extracción de información social desde Twitter y análisis mediante Hadoop.
  - Autor: Cristian Caballero Montiel
  - Tutores: Daniel Higuero Alonso-Mardones y Juan Manuel Tirado Martín
  - http://e-archivo.uc3m.es/handle/10016/16784

- Adaptation, Deployment and Evaluation of a Railway Simulator in Cloud Environments
  - Autora: Silvina Caíno Lores
  - Tutor: Alberto García Fernández

Agradecimientos

- Por último pero no por ello menos importante, agradecer al personal del Laboratorio del Departamento de Informática todos los comentarios y sugerencias para esta presentación.

Diseño de Sistemas Distribuidos

Máster en Ciencia y Tecnología Informática
Curso 2018-2019

Sistemas escalables en entornos distribuidos

Alejandro Calderón Mateos & Óscar Pérez Alonso
acaldero@inf.uc3m.es oscar@lab.inf.uc3m.es