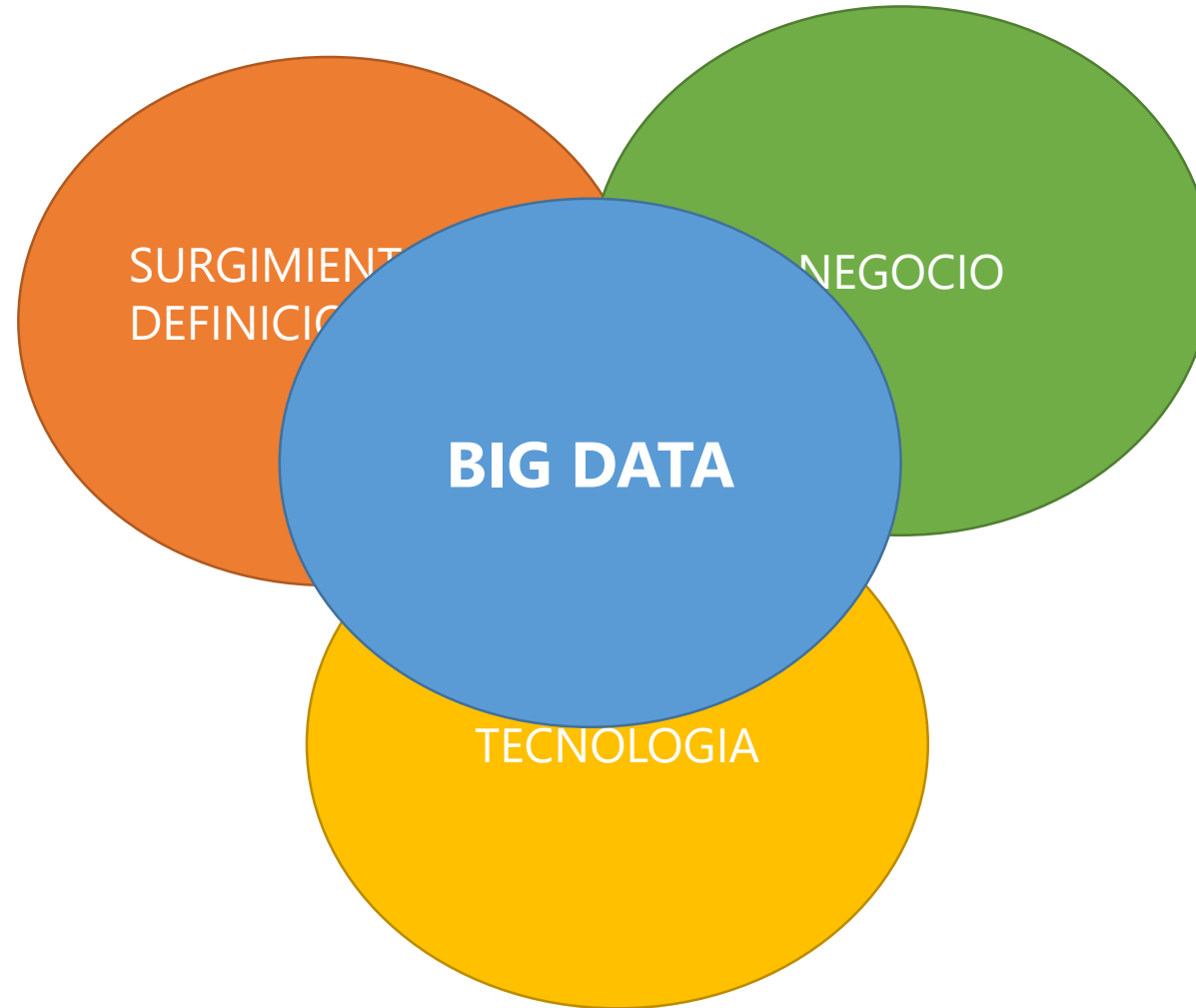


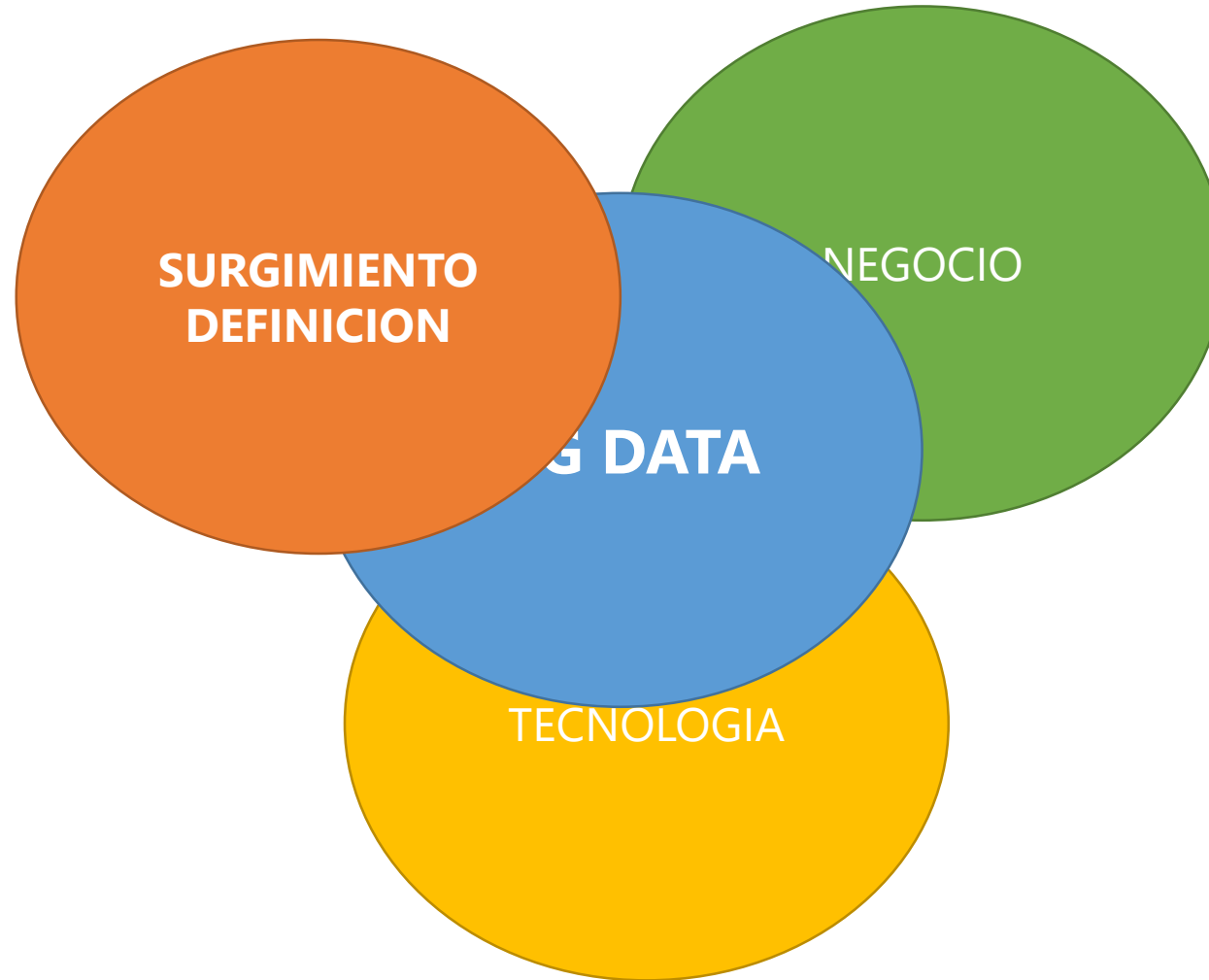
# Big Data

Tecnología e Innovación en la Industria

# BIG DATA



# DEFINIENDO BIG DATA



# DEFINIENDO BIG DATA

## SURGIMIENTO DE BIG DATA

- 1991 Surgimiento de Internet.
- 1996 El precio de **almacenamiento digital** cae. Es **más rentable que el papel**.
- 1997 **Google** lanza su motor de búsqueda.
- 1999 Primeras ideas de **Big Data**  
El primer uso del término Big Data **en un trabajo académico**: "Visually Exploring Gigabyte Datasets in Realtime (ACM)".  
El primer uso del concepto "**Internet of Things**" (Internet de las Cosas) en una presentación de negocios de Kevin Ashton para Procter and Gamble.
- 2001 Las tres "V" definida por Doug Laney para **Big Data**: "**Volumen, Velocity, Variety**".
- 2004 Se funda **Facebook** en Cambridge, Massachusetts.
- 2005 La **Web 2.0** aumenta el volumen de datos.
- 2006 Se funda **Twitter** en San Francisco, California.
- 2007 Emerge el uso actual del término "**Big Data**"  
Steve Jobs, Apple, lanza el **Iphone**.

# DEFINIENDO BIG DATA

## SURGIMIENTO DE BIG DATA

- 2009 Una **compañía promedio estadounidense** con más de **1.000 empleados**, **almacena** más de **200 terabytes** de datos de acuerdo al informe sobre Big Data: "The Next Frontier for Innovation, Competition and Productivity" realizado por McKinsey Global Institute.
- 2010 Eric Schmidt, **presidente ejecutivo de Google**, dice en una conferencia que la cantidad de datos que ahora se están **creando cada dos días**, **es mayor** que la creada **desde el comienzo de la civilización humana** hasta el año 2003.
- 2014 El **uso de Internet móvil supera** a las **computadoras de escritorio** por primera vez. El **88% de los ejecutivos** que respondieron a una encuesta internacional realizada por GE dice que el **análisis de grandes volúmenes de datos es una prioridad**.

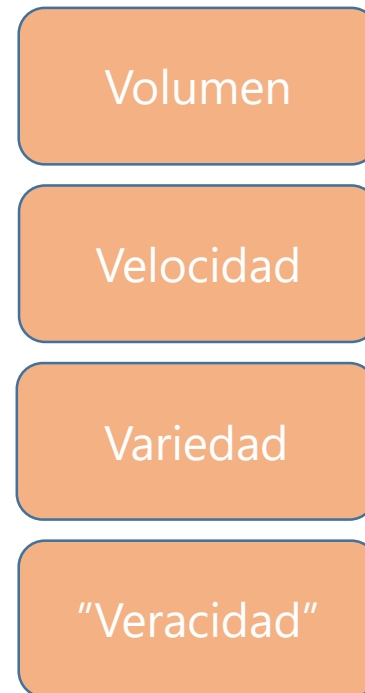
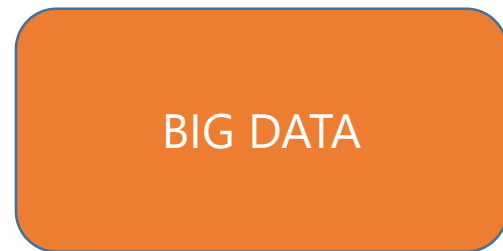
# DEFINIENDO BIG DATA

## Big Data Una Primer Definición

*"Volumen masivo de datos, tanto estructurados como no-estructurados, los cuales son demasiado grandes y difíciles de procesar con las bases de datos y el software tradicionales."* (ONU, 2012)

# DEFINIENDO BIG DATA

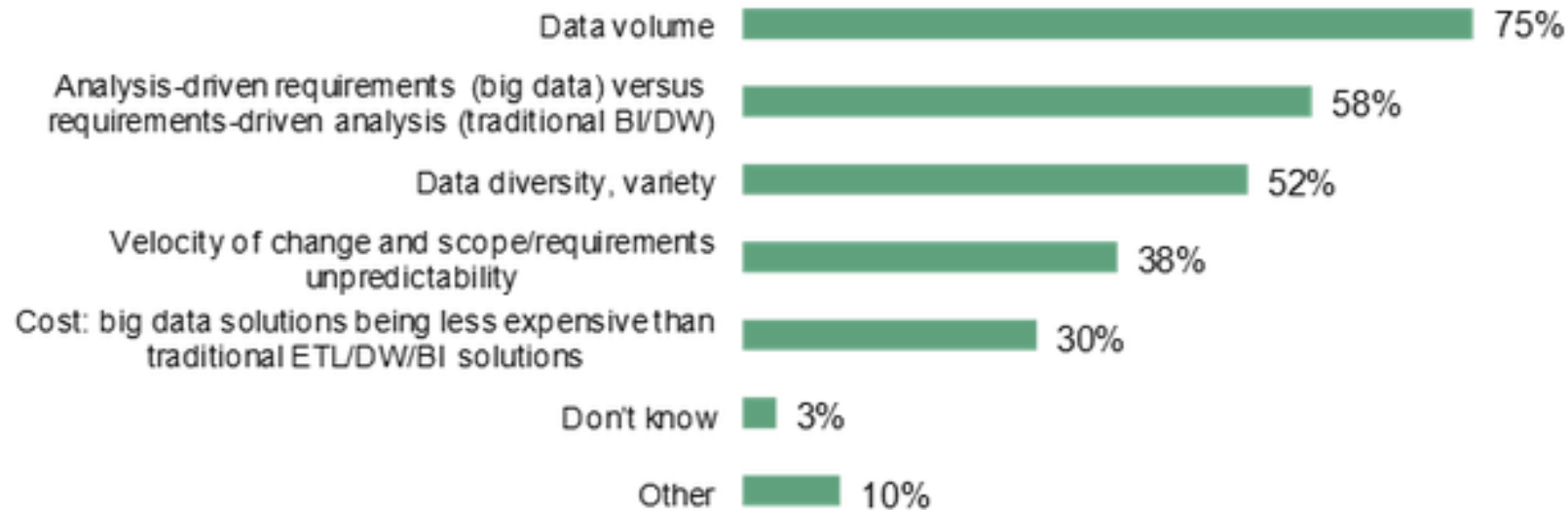
Big Data es el sector de IT que hace referencia a *grandes conjuntos de datos* que por la *velocidad* a la que se generan, la capacidad para tratarlos y los *múltiples formatos y fuentes*, es necesario procesarlos con mecanismos distintos a los tradicionales.



# DEFINIENDO BIG DATA

## ¿ CUÁLES DE LAS 4 Vs TIENEN MAYOR INFLUENCIA ?

En ambientes tradicionales de BI y DW primero se generan los requerimientos y luego las aplicaciones. Dicho de otra forma, los requerimientos direccionan las aplicaciones. En Big Data es al revés, ya que se utiliza la exploración de datos libre para generar hipótesis para encontrar un patrón



El costo es un factor en muchos casos. Las tecnologías utilizadas en Big Data son más económicas que las tradicionales.

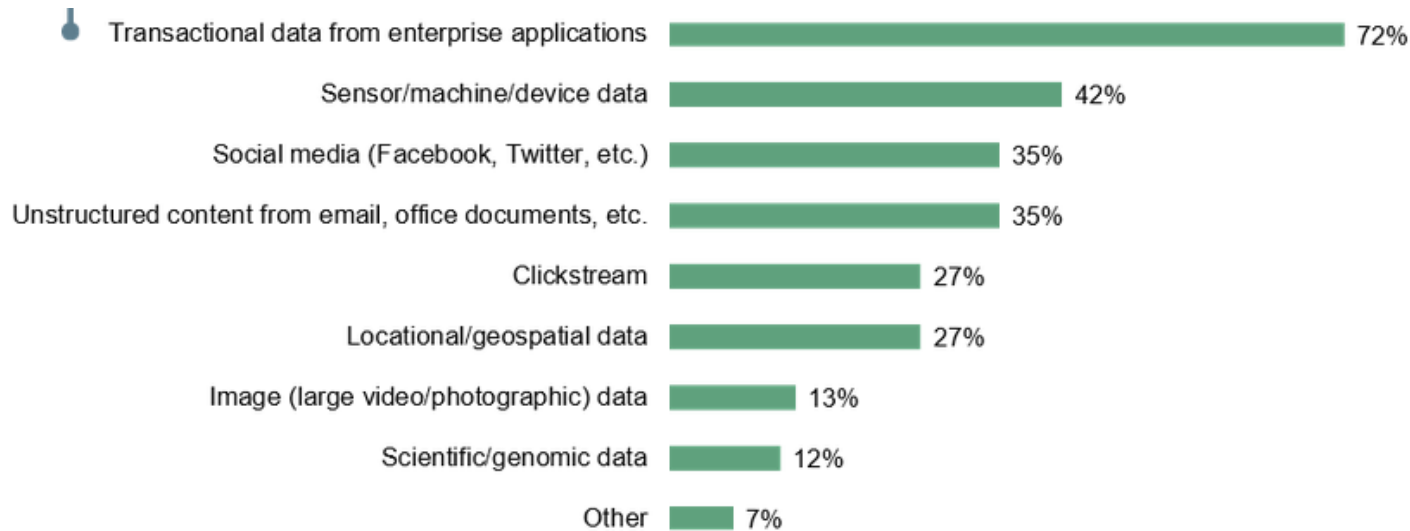
fuelle: forrester research. global big data survey



# DEFINIENDO BIG DATA

## TIPOS DE DATOS QUE SE ANALIZAN A TRAVÉS DE BIG DATA

Si bien pareciera que Big Data se utiliza fundamentalmente para datos puros obtenidos de redes sociales, sensores, tráfico web la realidad es que más allá de lo que se supone, las empresas utilizan Big Data para operar en datos operacionales.



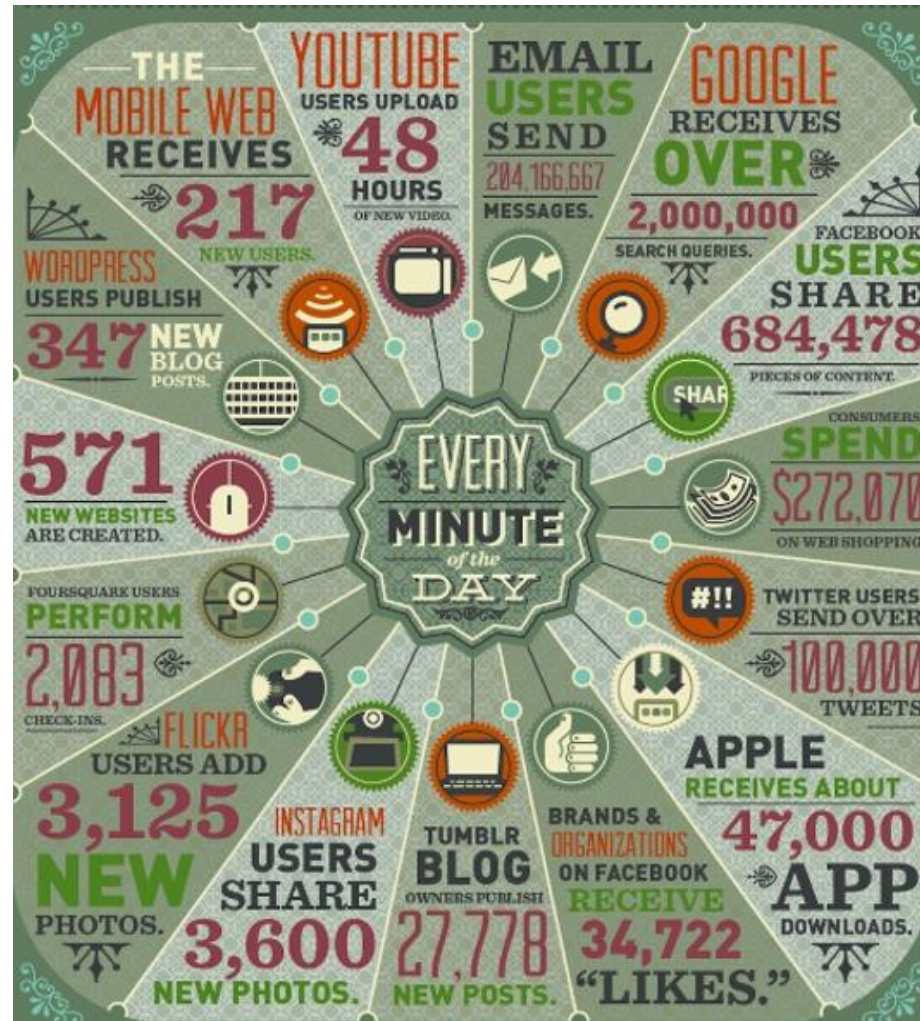
fuelle: forrester research. global big data survey

# DEFINIENDO BIG DATA



# DEFINIENDO BIG DATA

*Año 2012*



Por cada minuto del día

YouTube 48 hs. de Video

Google +2,000,000 Búsquedas

Facebook 684.478 Piezas Compartidas

Twitter 100,000 tweets

Apple 47000 Apps Download

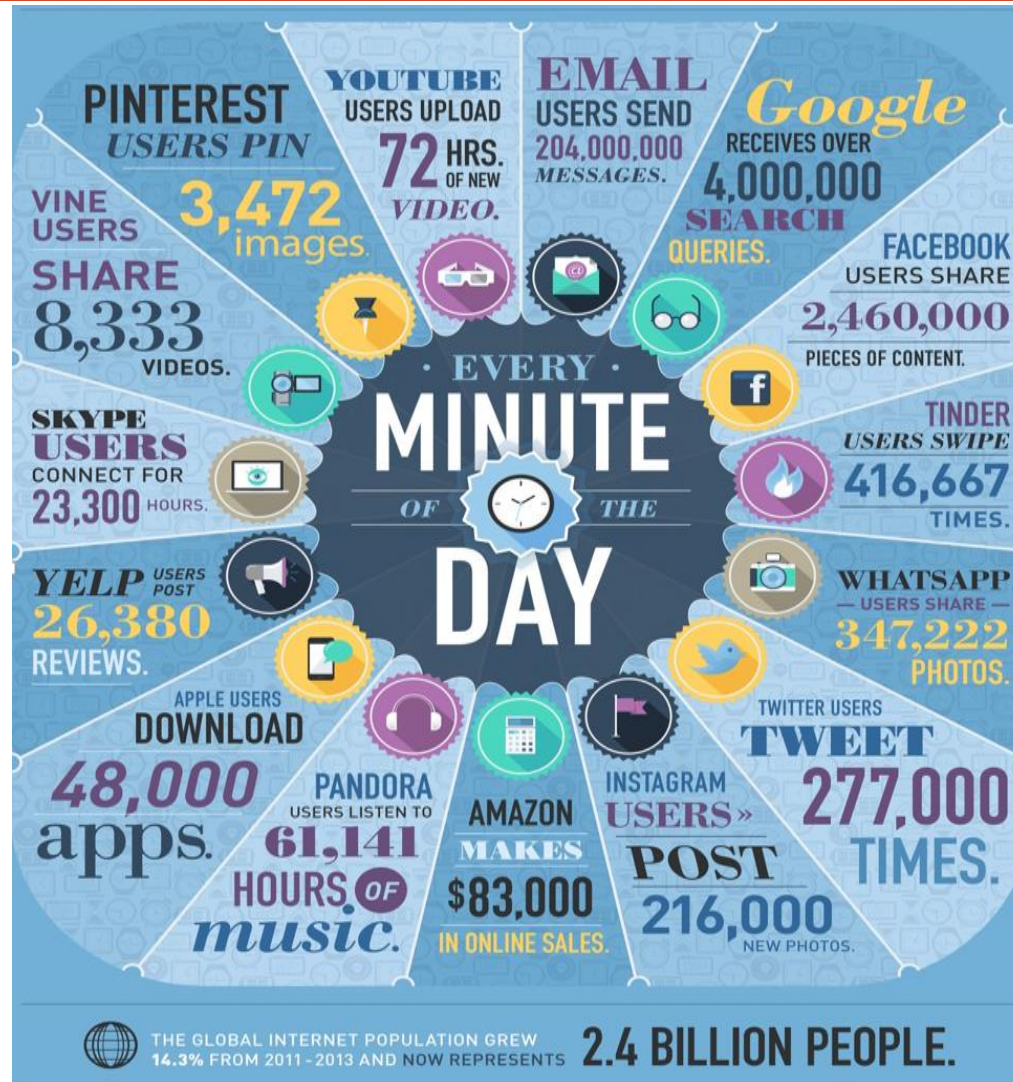
**Población Total de Internet**

**2.100.000.000 de personas**



# DEFINIENDO BIG DATA

*Año 2013*



Por cada minuto del día

YouTube	72 hs. de Video
Google	+4,000,000 Búsquedas
Facebook	2,460.000 Piezas Compartidas
Twitter	277,000 tweets
Apple	48,000 Apps Download
Whatsapp	347,222 Photos Sended
Tinder	416,667 Users Swipe

**Población Total de Internet**

**2.400.000.000 de personas**

# DEFINIENDO BIG DATA

*Año 2015*



Por cada minuto del día

YouTube 300 hs. de Video  
Facebook 4,166.667 Posts  
Twitter 347,222 tweets  
Apple 51,000 Apps Download  
Whatsapp 347,222 Photos  
Uber 694 pasajeros  
Tinder 590,278 Users Swipe  
SnapChat 284,722 Snaps

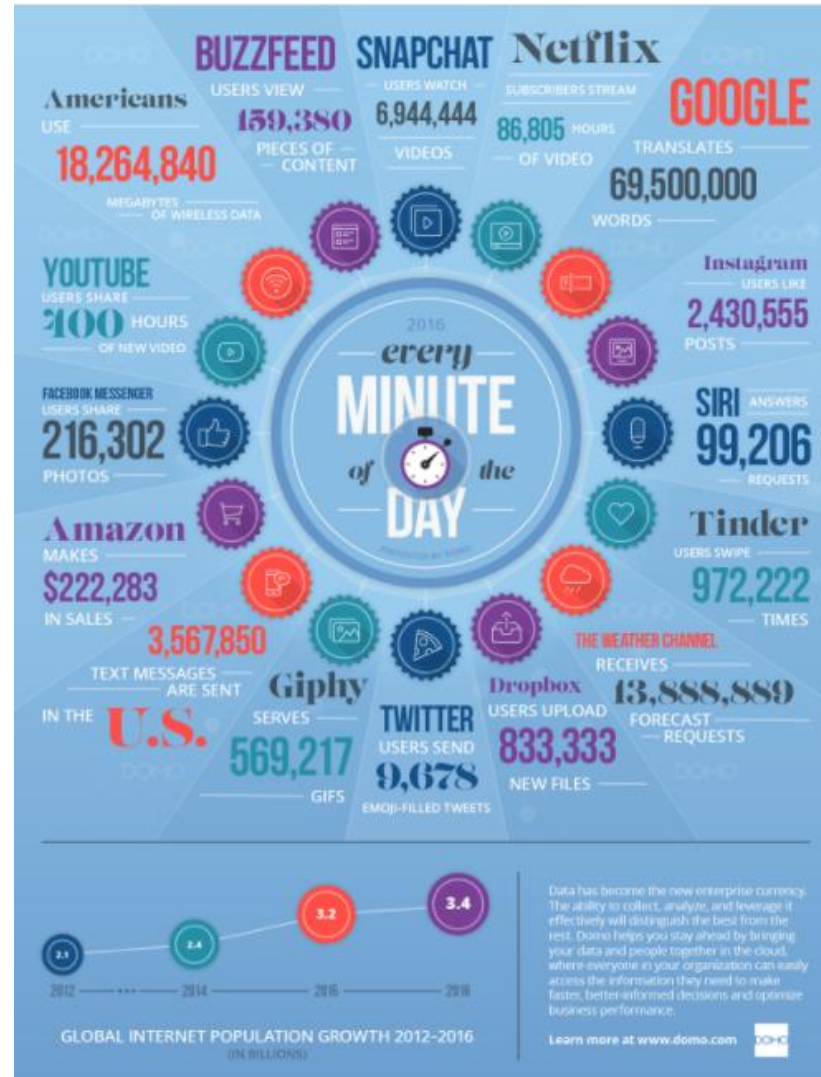
**Población Total de Internet**

**3.200.000.000 de personas**



# DEFINIENDO BIG DATA

*Año 2016*



Por cada minuto del día

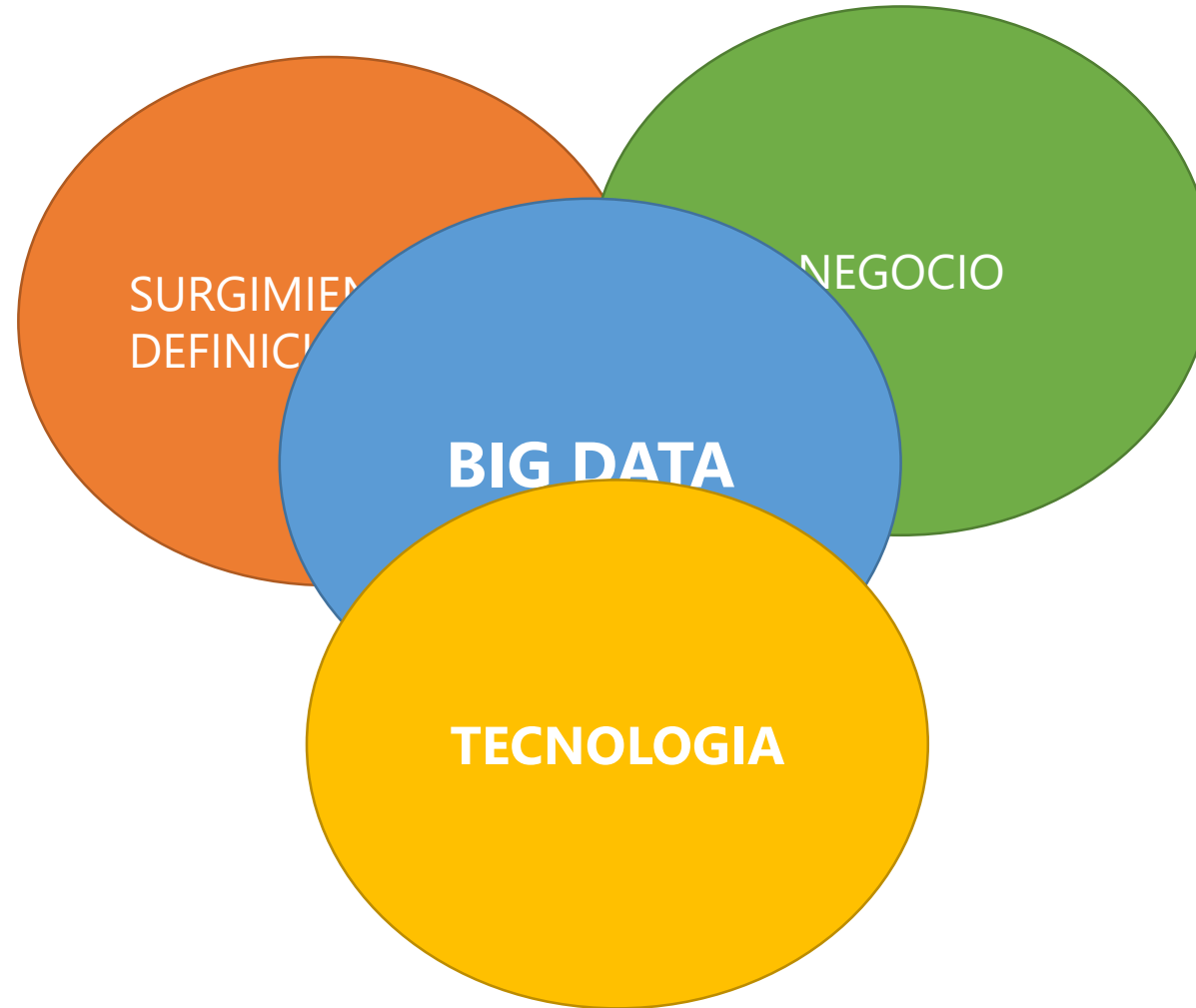
YouTube 400 hs. de Video  
Google 69.500.000 palabras traducidas  
Netflix 86,805 Horas de Video  
Facebook 4,166.667 User share  
Siri 99,206 Requests  
Whatsapp 347,222 Photos  
Tinder 972,222 Users Swipe  
SnapChat 6,944,444 Videos Watched  
TWC 13,888,889 Pronósticos

**Población Total de Internet**

**3.400.000.000 de personas**



# BIG DATA



# IMPLEMENTANDO BIG DATA

## PRINCIPALES CAMBIOS QUE SE PRODUJERON EN LA TECNOLOGÍA Y EN LOS ÚLTIMOS 15 AÑOS

- MASIFICACIÓN USO DE INTERNET
- SURGIMIENTO DE LAS REDES SOCIALES
- CRECIMIENTO EXPONENCIAL DE DISPOSITIVOS MÓVILES
- INTERFACES DE USUARIO MAS SIMPLES E INTUITIVAS
- CAMBIOS EN LAS FORMAS DE PROCESAMIENTO
- FUERTE BAJA EN LOS COSTOS DE ALMACENAMIENTO

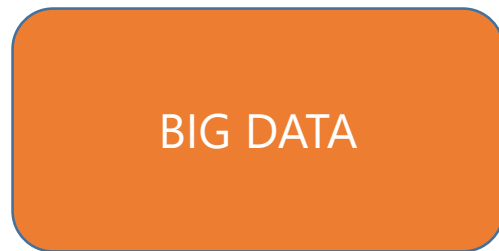
CADA DÍA CREAMOS 2,5  
QUINTILLONES DE BYTES DE  
DATOS. (2,5 Exabytes)

EL 90% DE LOS DATOS DEL  
MUNDO DE HOY SE  
GENERARON EN LOS ÚLTIMOS  
2 AÑOS



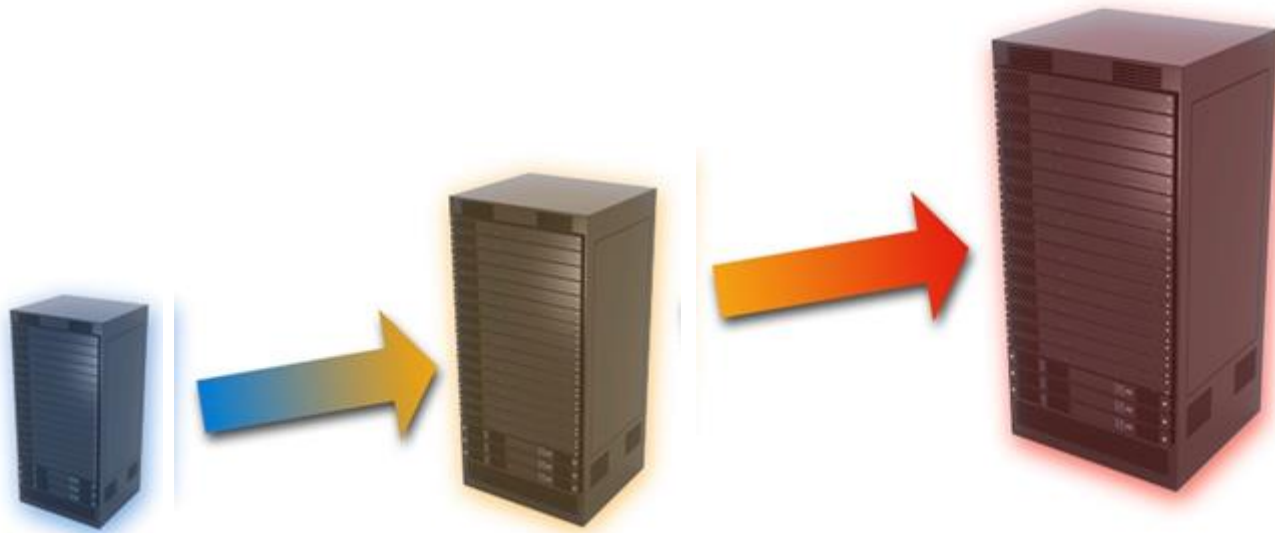
# IMPLEMENTANDO BIG DATA

Big Data es el sector de IT que hace referencia a *grandes conjuntos de datos* que por la *velocidad* a la que se generan, la capacidad para tratarlos y los *múltiples formatos y fuentes*, es necesario procesarlos con mecanismos distintos a los tradicionales.



# IMPLEMENTANDO BIG DATA

## ESCALAMIENTO



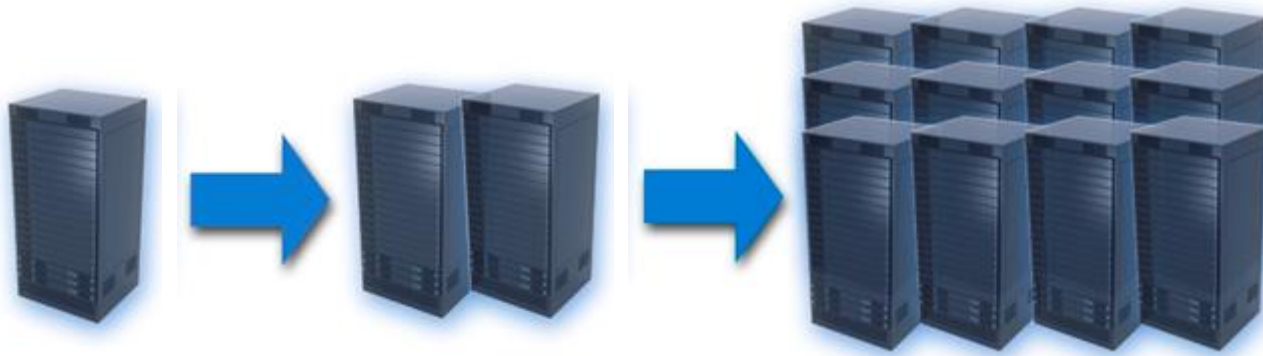
### Escalamiento Vertical

- Escalamiento dentro de un mismo servidor.
- Implica incrementar la capacidad de un Servidor agregando más recursos de CPU, memoria y de almacenamiento.

Escalamiento Vertical

# IMPLEMENTANDO BIG DATA

## ESCALAMIENTO



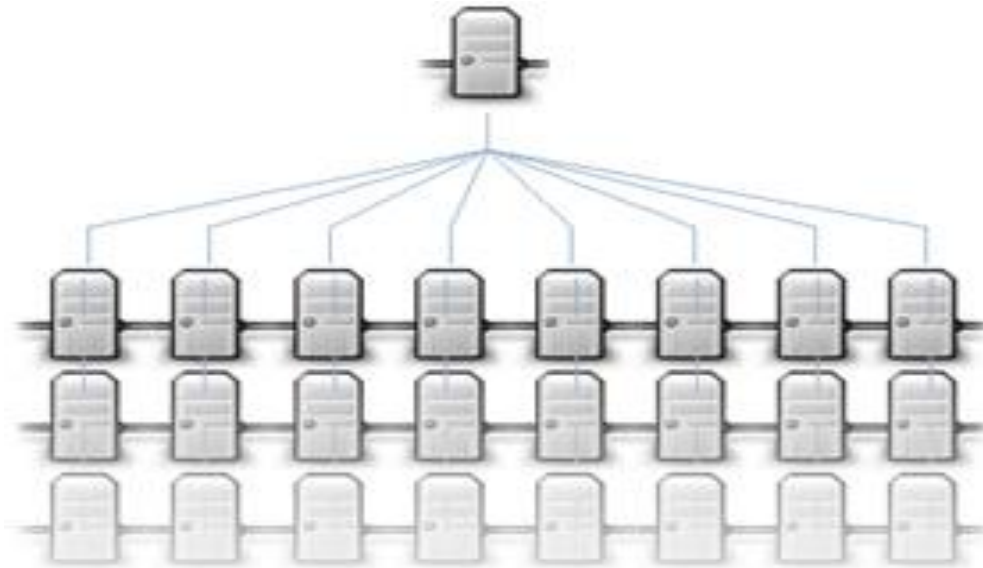
### Escalamiento Horizontal

- Escalamiento en varios servidores.
- Cluster de Servidores.
- Replicación de Datos.
- Particionamiento de Datos.
- Procesamiento Paralelo.

Escalamiento Horizontal

# IMPLEMENTANDO BIG DATA

## Cluster



Google

amazon

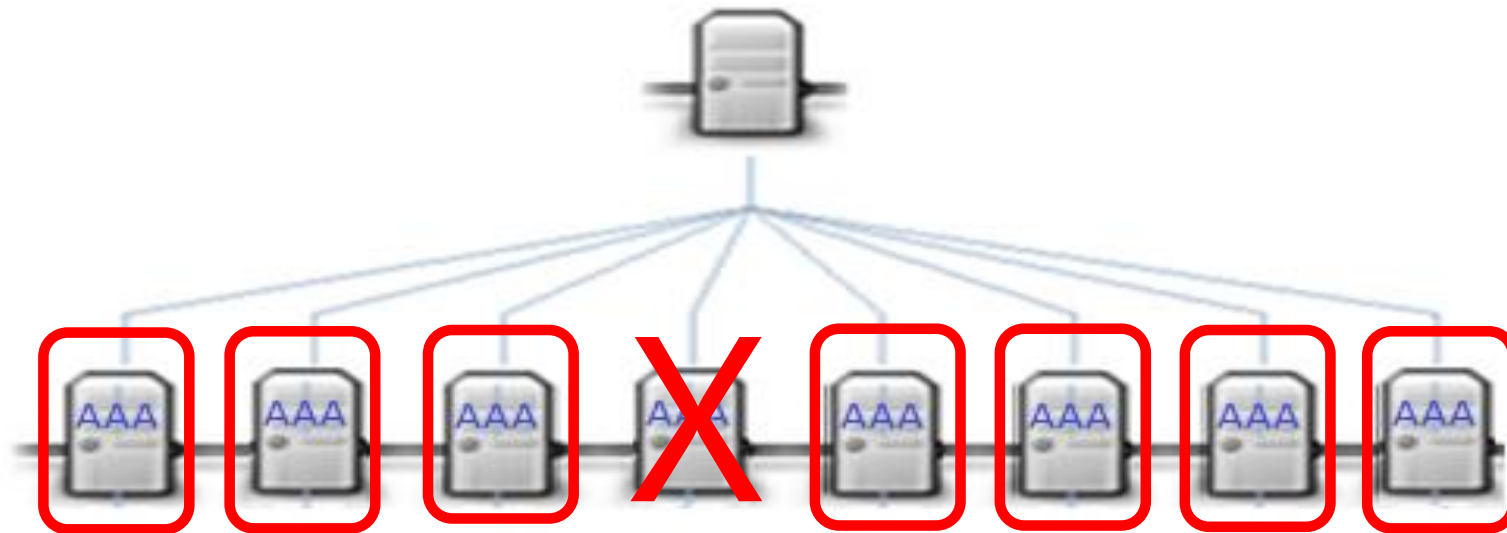
facebook

Grupo de servidores independientes interconectados a través de una red dedicada que trabajan como un único recurso de procesamiento

# IMPLEMENTANDO BIG DATA

## ALTA DISPONIBILIDAD Y TOLERANCIA A FALLOS

- Aplicaciones 7 x 24.
- Aplicaciones de Misión Crítica.

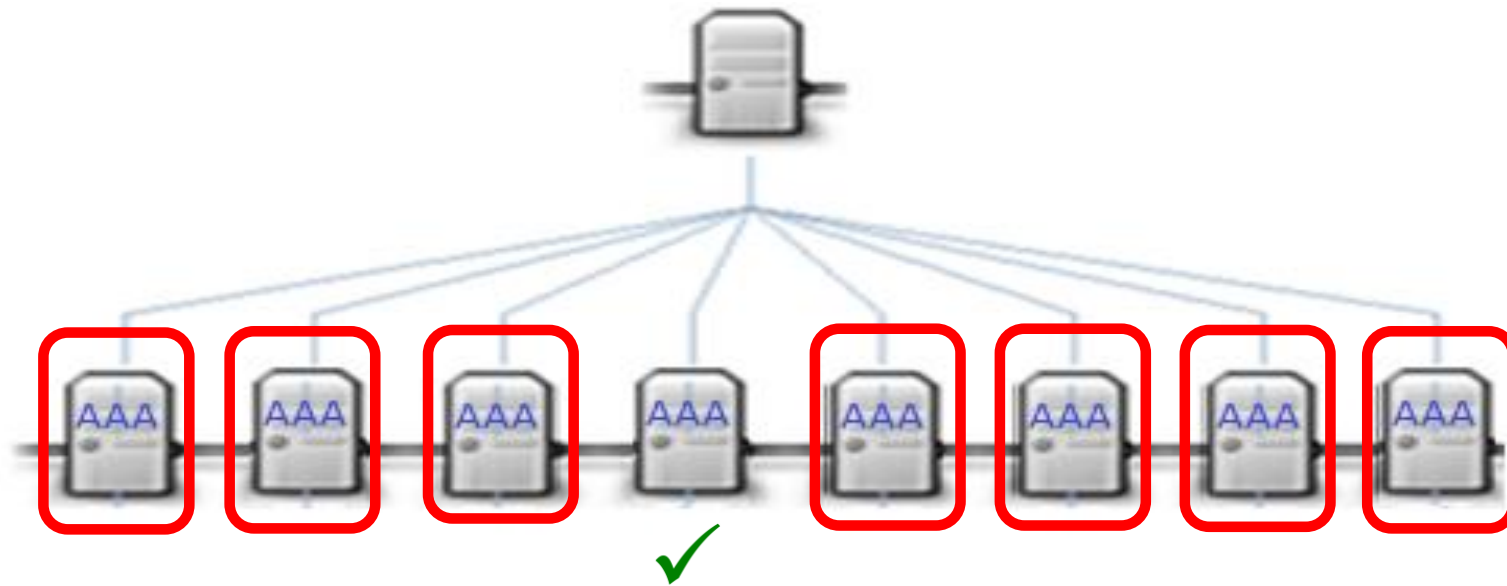


- Replicación

# IMPLEMENTANDO BIG DATA

## ALTA DISPONIBILIDAD Y TOLERANCIA A FALLOS

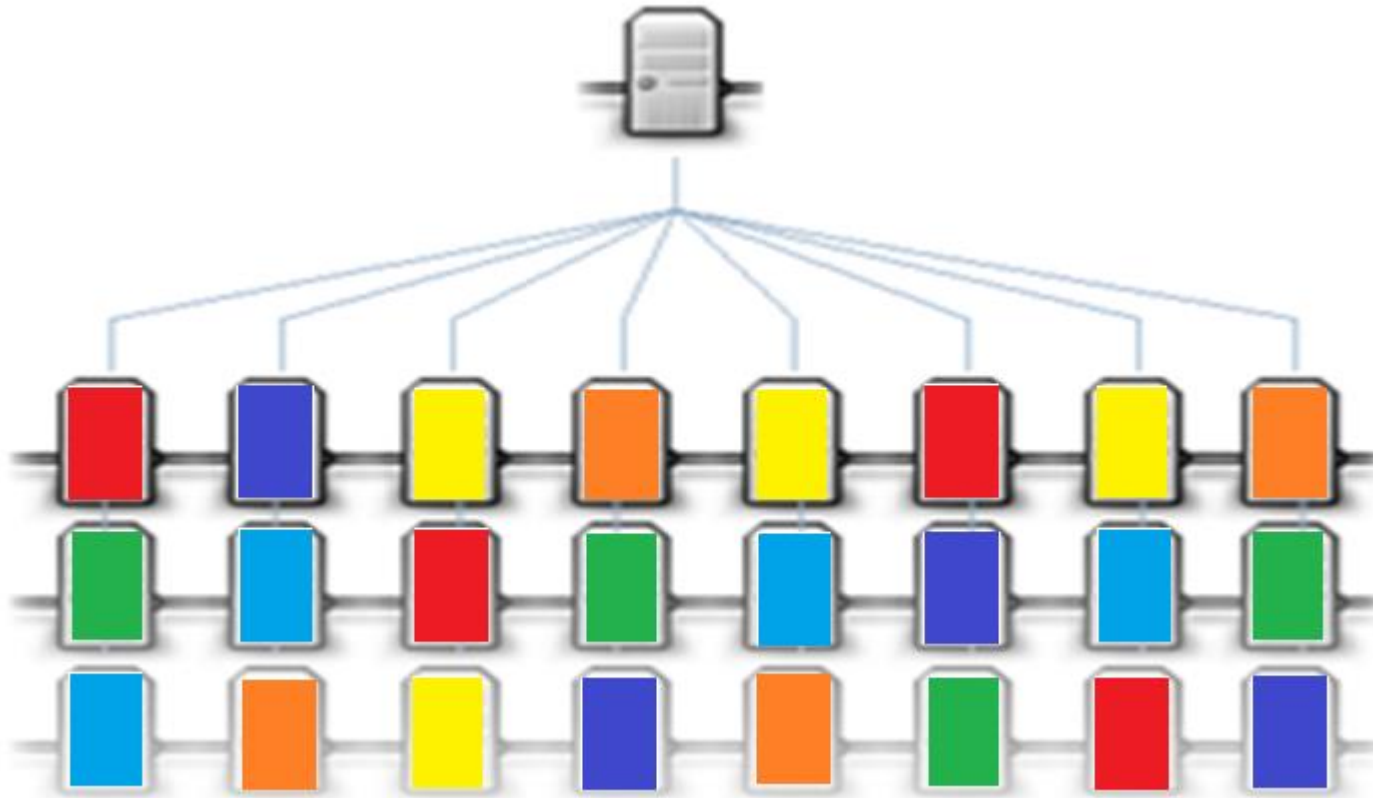
- Aplicaciones 7 x 24.
- Aplicaciones de Misión Crítica.



- Replicación

# IMPLEMENTANDO BIG DATA

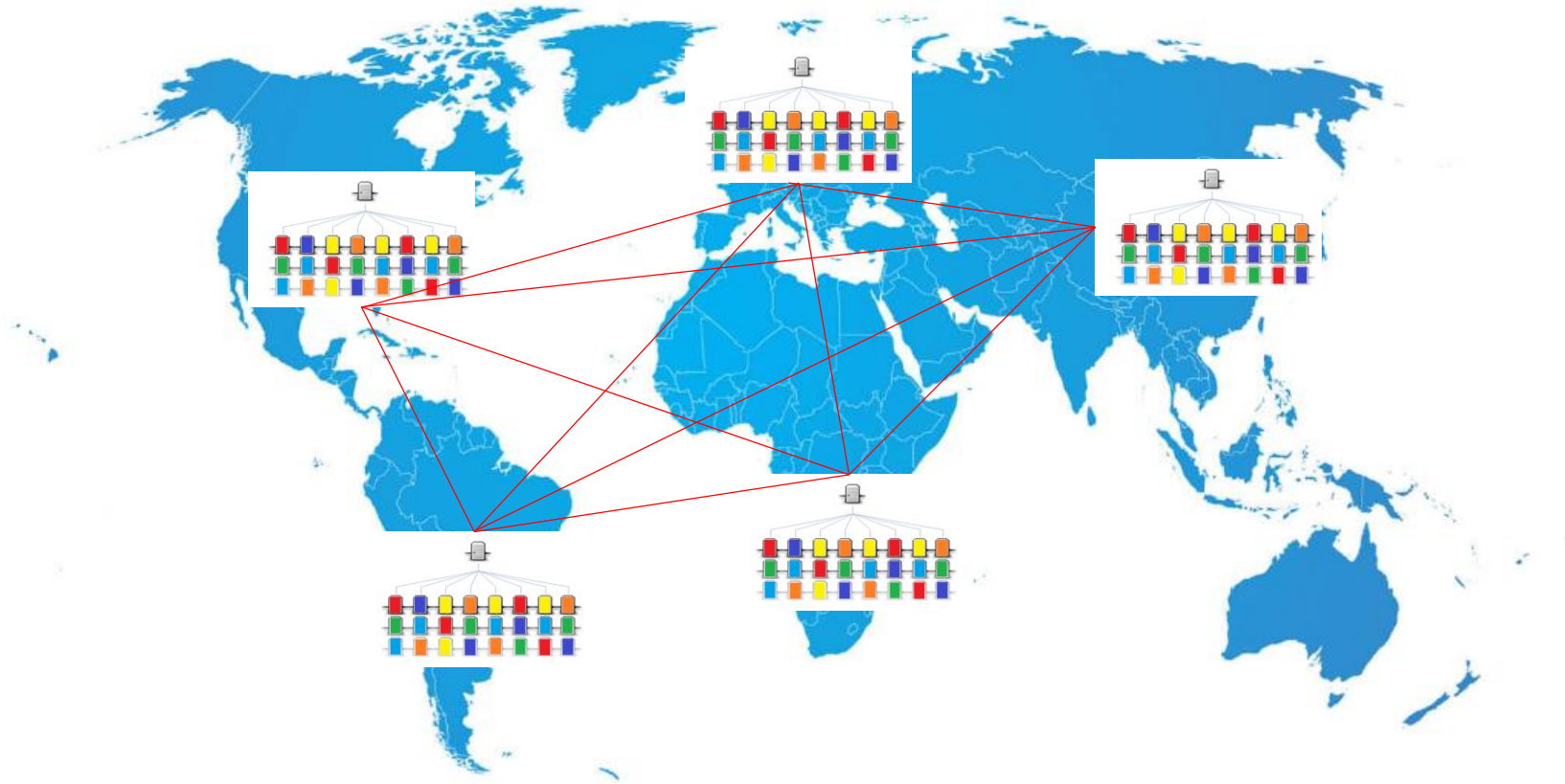
## PARTICIONAMIENTO DE DATOS



- Un solo Servidor no soporta almacenar la totalidad de los datos.
- Se deben particionar los datos en múltiples Servidores del Cluster.
- Además los datos se encuentran replicados.

# IMPLEMENTANDO BIG DATA

## PARTICIONAMIENTO DE DATOS

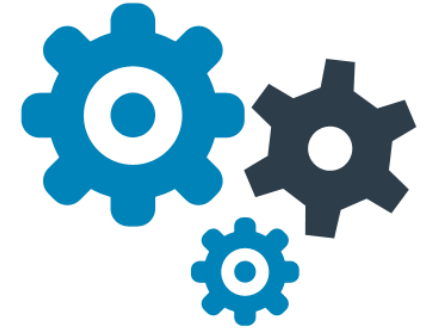


- Necesidad de distribución geográfica de datos.



# IMPLEMENTANDO BIG DATA

## PROCESAMIENTO PARALELO



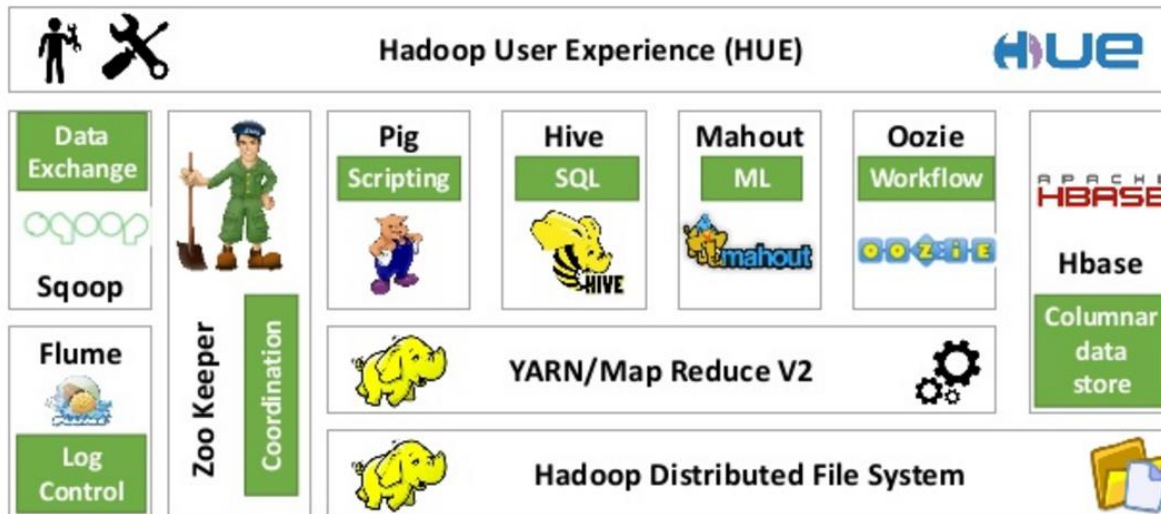
- Varios servidores procesan un mismo programa de forma simultánea para resolver un determinado problema.

# IMPLEMENTANDO BIG DATA

## TECNOLOGÍAS EXISTENTES EN EL MERCADO

# IMPLEMENTANDO BIG DATA

## HADOOP

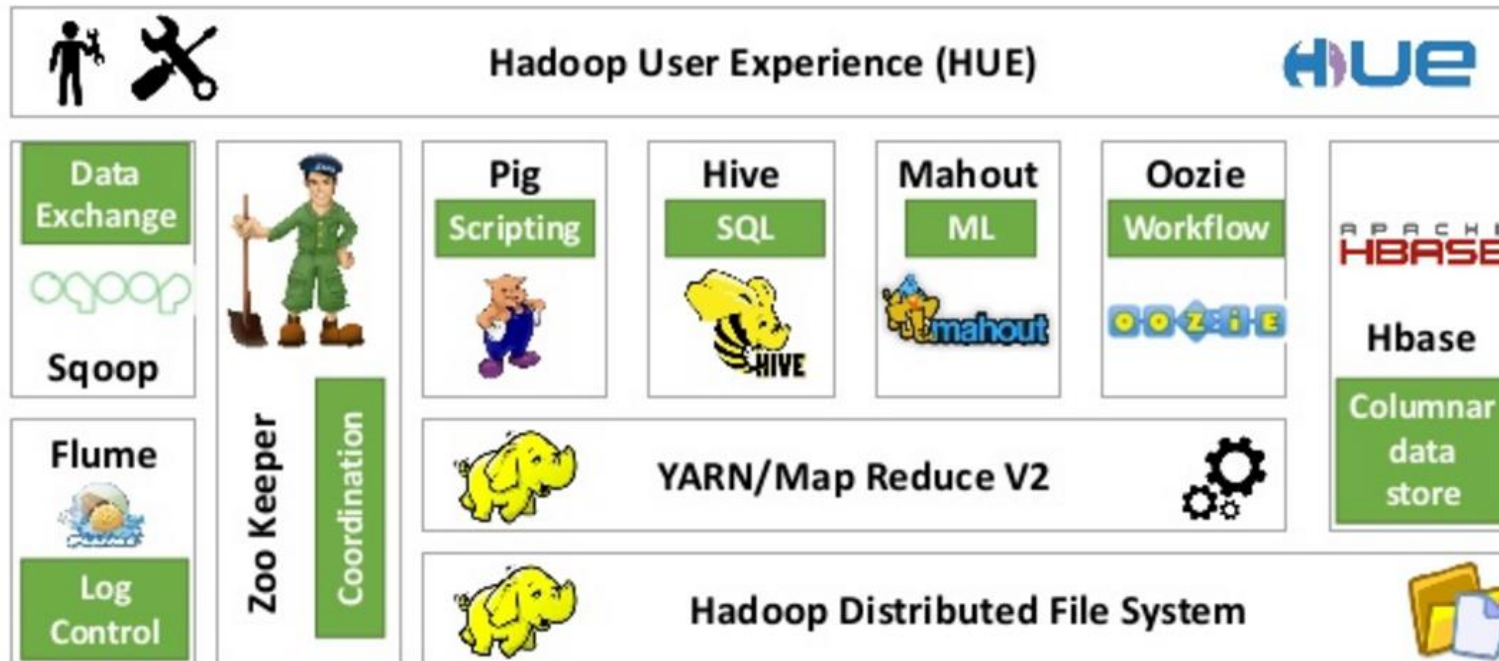


¿ Qué es Hadoop ?

**Hadoop** es un **sistema de código abierto** que se utiliza para **almacenar, procesar en paralelo y analizar grandes volúmenes de datos**.

# IMPLEMENTANDO BIG DATA

## HADOOP - Stack



### HDFS

Sistema de Archivos Distribuidos.

### YARN

Gestor de Recursos y Contenedor de Motores de Procesamiento Paralelo.

### Map Reduce / Tez / Spark

Motores de Procesamiento Paralelo.

### Pig / Hive

Aplicaciones para desarrollo de consultas y programas para procesamiento distribuido.

### Hbase

Base de Datos NoSQL Columnar.

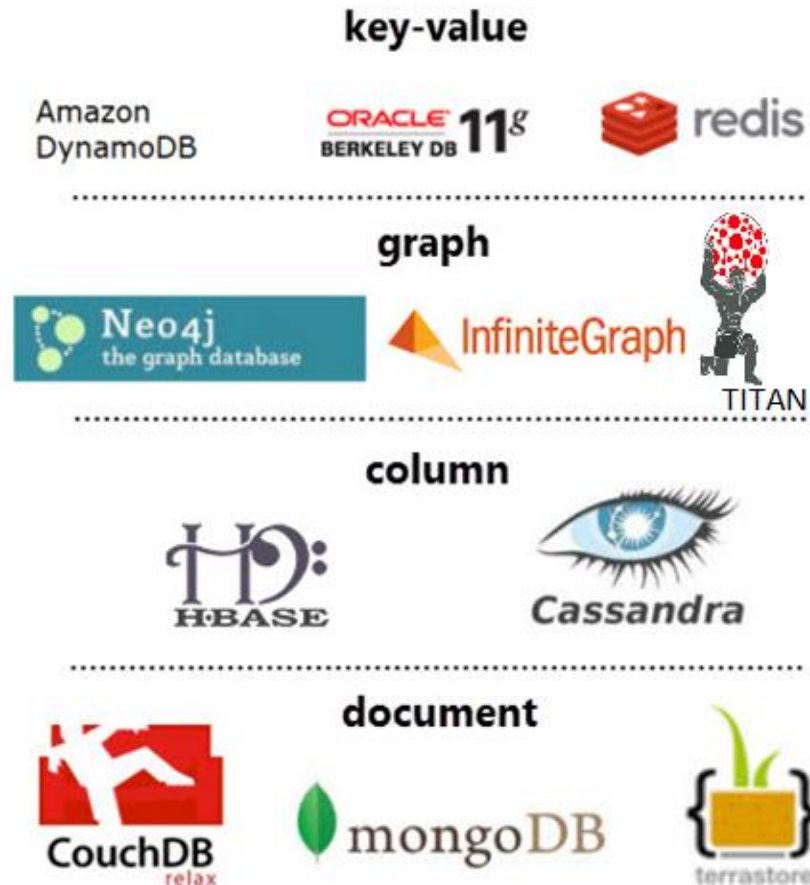
# IMPLEMENTANDO BIG DATA

## HADOOP - Distribuciones



# IMPLEMENTANDO BIG DATA

## BASES DE DATOS NO-SQL



### ¿ Qué es NoSQL ?

Sistemas de gestión de bases de datos que difieren del modelo clásico de bases de datos relacionales: no usan SQL como lenguaje de consulta, los datos almacenados no requieren estructuras fijas como tablas, no garantizan consistencia plena y escalan horizontalmente.

### ¿ Qué es la Pesistencia Políglota ?

Utilizar dentro de un mismo ambiente o aplicación un conjunto de bases de datos, que colabora, cada una en lo que es más importante.

# IMPLEMENTANDO BIG DATA

## Persistencia Políglota

**Búsquedas Performantes  
sobre Catálogo de  
Productos**



**Información Distribuida  
Geográficamente  
Profile de usuarios y  
Documentación de  
Productos con  
Info no estructurada**



**Caché de Sesiones  
Lockeos Distribuidos**

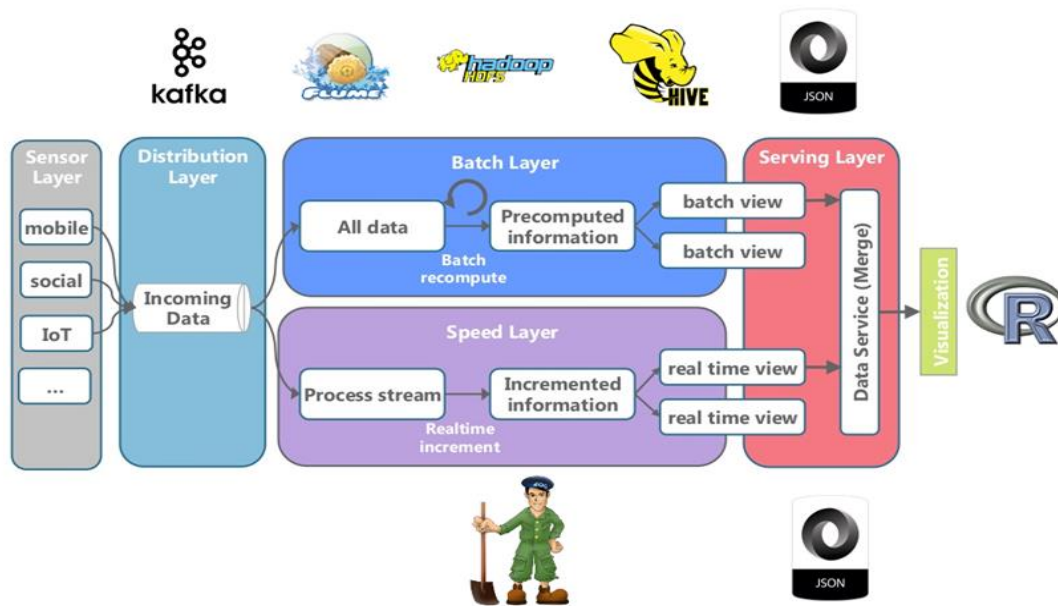


**Transacciones  
Económicas.**



# IMPLEMENTANDO BIG DATA

## ARQUITECTURA LAMBDA



**Arquitectura Lambda** es una arquitectura diseñada para manejar cantidades masivas de datos aprovechando las ventajas de dos métodos de procesamiento, procesamiento por lotes (batch) y procesamiento de flujo de datos (streaming).

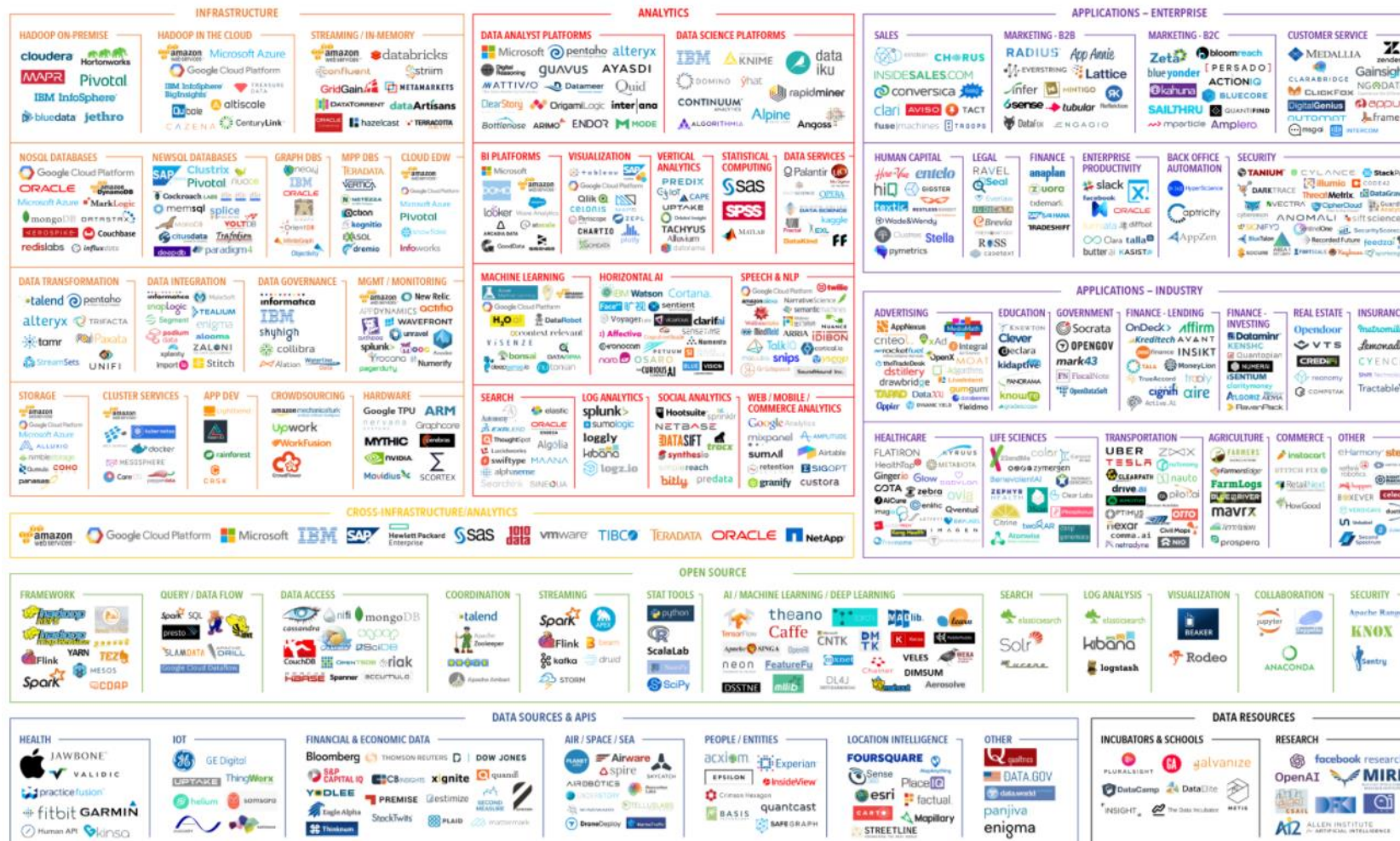
Proporcionar vistas completas y precisas de los datos del lote, y al mismo tiempo vistas en tiempo real de datos en movimiento.

Las Salidas pueden unirse antes de la presentación.

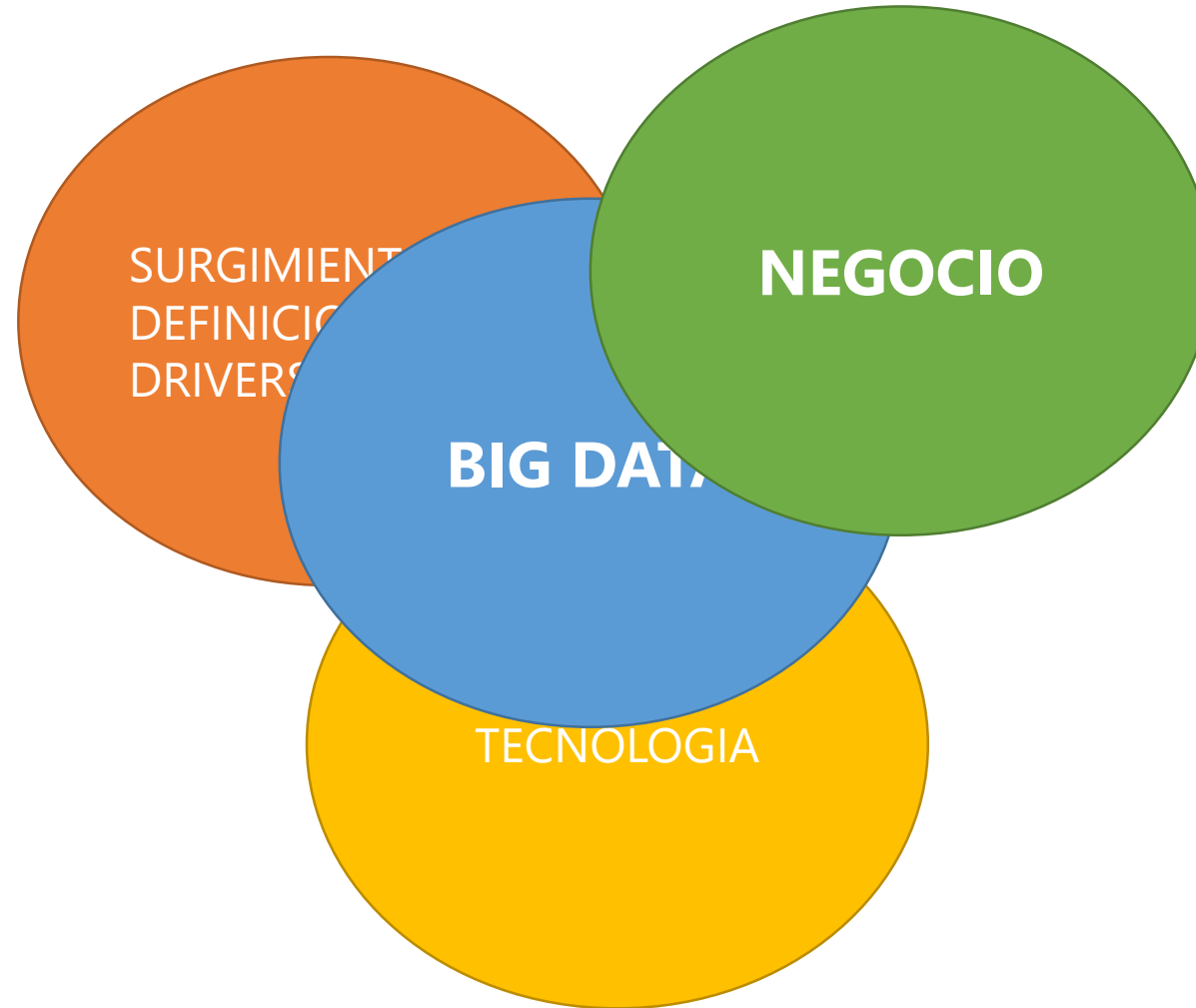


# IMPLEMENTANDO BIG DATA

## BIG DATA LANDSCAPE 2017

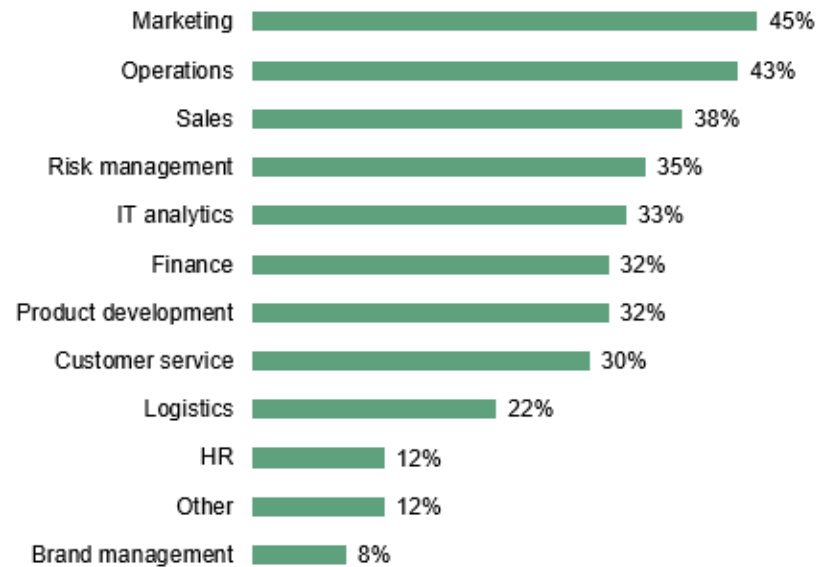


# BIG DATA



# ÁREAS DE NEGOCIO QUE LIDERAN EL USO DE BIG DATA.

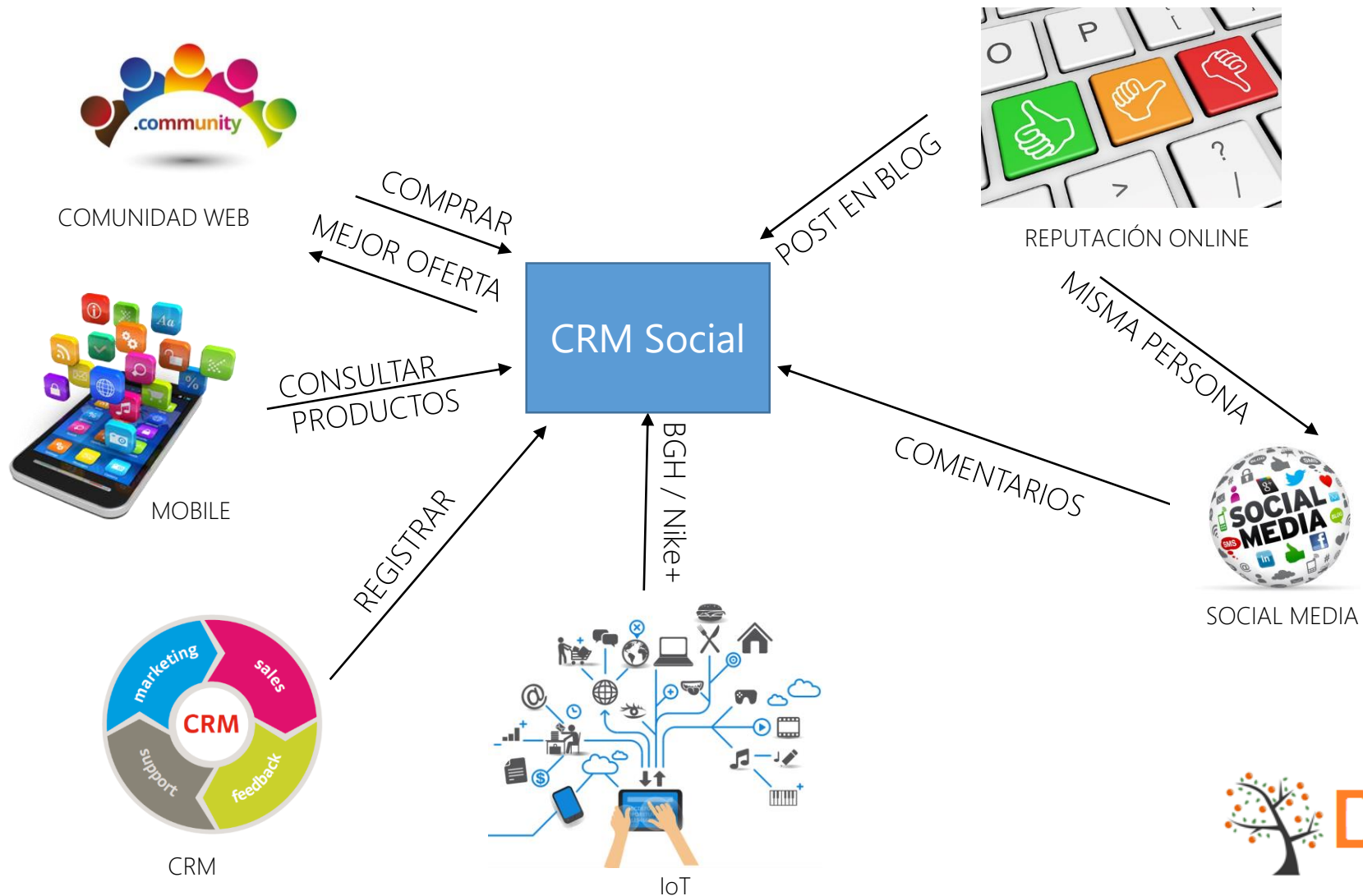
Big Data resuelve muchos problemas en una gran variedad de dominios de negocio, pero Marketing, Operaciones y Ventas lideran el ranking.



Gestión de Riesgos y Finanzas són áreas de crecimiento

fuelle: forrester research. global big data survey

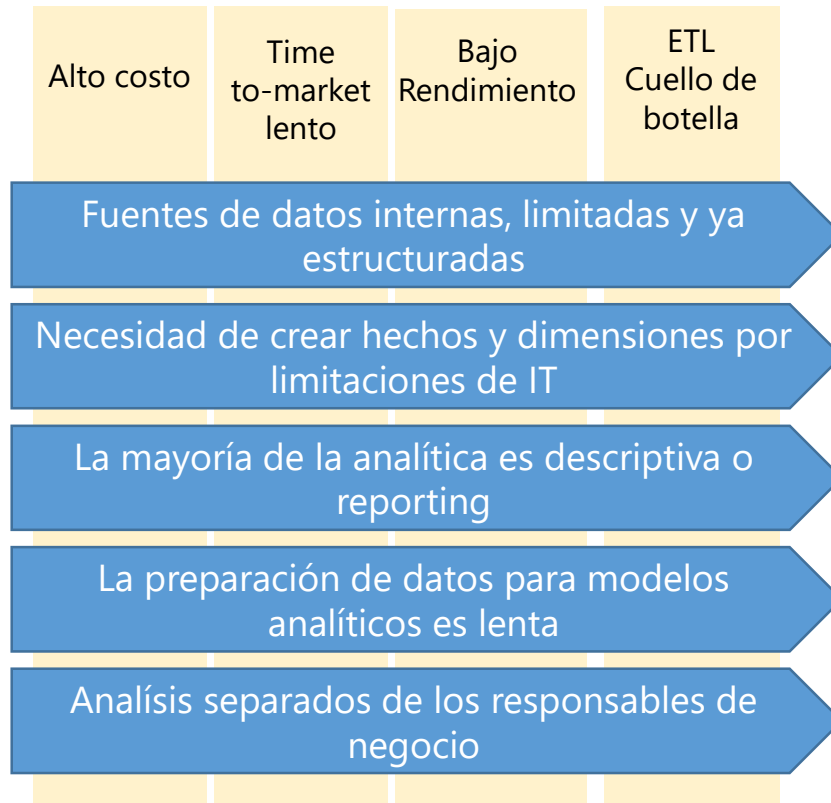
# BIG DATA Y LOS NEGOCIOS



Break  
En 15 minutos  
volvemos...

# BIG DATA Y LOS NEGOCIOS

## MODELO TRADICIONAL de BI



## ANALÍTICA 3.0

Acceso total a la información.  
"High Resolution View"

Dar respuesta al negocio, sin  
limitar la visión de negocio

Modelización predictiva

Paralelización de procesos

Perfiles avanzados: Data  
Scientists

# BIG DATA Y LOS NEGOCIOS

## QUE PUEDE LOGRAR LA EMPRESA

- **Predicciones:** Predecir patrones de gustos y necesidades de los clientes o potenciales clientes.
- **Innovación:** Innovar para crear una experiencia a medida para sus clientes.
- **Mejorar los esfuerzos de marketing :** Aplicar inteligencia al momento de realizar la gestión de las campañas de marketing. Recomendación de contenidos.
- **Velocidad:** Detectar y analizar On-line comportamiento, gustos y necesidades

# BIG DATA Y LOS NEGOCIOS

## QUE PUEDE LOGRAR LA EMPRESA

- **Tecnología:** Recopilar información multiplataforma analizando el uso de los diferentes dispositivos de visualización para optimizar la publicidad y la oferta de productos
- **Retención y generación de clientes:** Proponer productos que sean los preferidos por los usuarios y generar nuevos clientes a través del análisis detallado.
- **Obtención de Información detallada de clientes:**
  - A través de recopilación de información pública (redes sociales, web, etc.), interna, financiera y agregada de los usuarios podemos determinar la ubicación de los usuarios con fines de consumo (geo-referenciación, información demográfica).
  - Realizar análisis de comportamiento de compras pasadas, sentiment analysis de redes sociales y datos de los CRM.







# IoT – Ingresando a una nueva era

La cantidad de dispositivos conectados supera la población mundial

Internet of Things “IoT” a Internet of Everything “IoE”.

El 99% de las cosas no están conectadas aún.

Para el 2020 se sumaran 2.5 billones de usuarios a la web.

Para el 2020 se sumaran 3.7 billones de dispositivos a la web.

# Evolución de conexiones de datos

Category	2013	2014	2015	2020
Automotive	96.0	189.6	372.3	3,511.1
Consumer	1,842.1	2,244.5	2.874.9	13,172.5
Generic Business	395.2	479.4	623.9	5,158.6
Vertical Business	698.7	836.5	1,009.4	3,164.4
<b>Total</b>	<b>3,032.0</b>	<b>3,750.0</b>	<b>4,880.6</b>	<b>25,006.6</b>

Source: Gartner

Unidad de Medida en Millones

<http://www.gartner.com/newsroom/id/2905717>



# Conclusión en esta nueva era

- Todo va a estar conectado
- Toda conexión tiene una voz
- Esa voz tiene un mensaje / información / historia
- El desafío es conectar esas historias, descubrir que tienen para decir e integrarlas a nuestros procesos de análisis de la compañía para conocer y ofrecer mejores productos / servicios a nuestros clientes.

# CASOS DE USO

## CASOS DE USO EN ARGENTINA Y LA REGIÓN

# CASOS DE USO

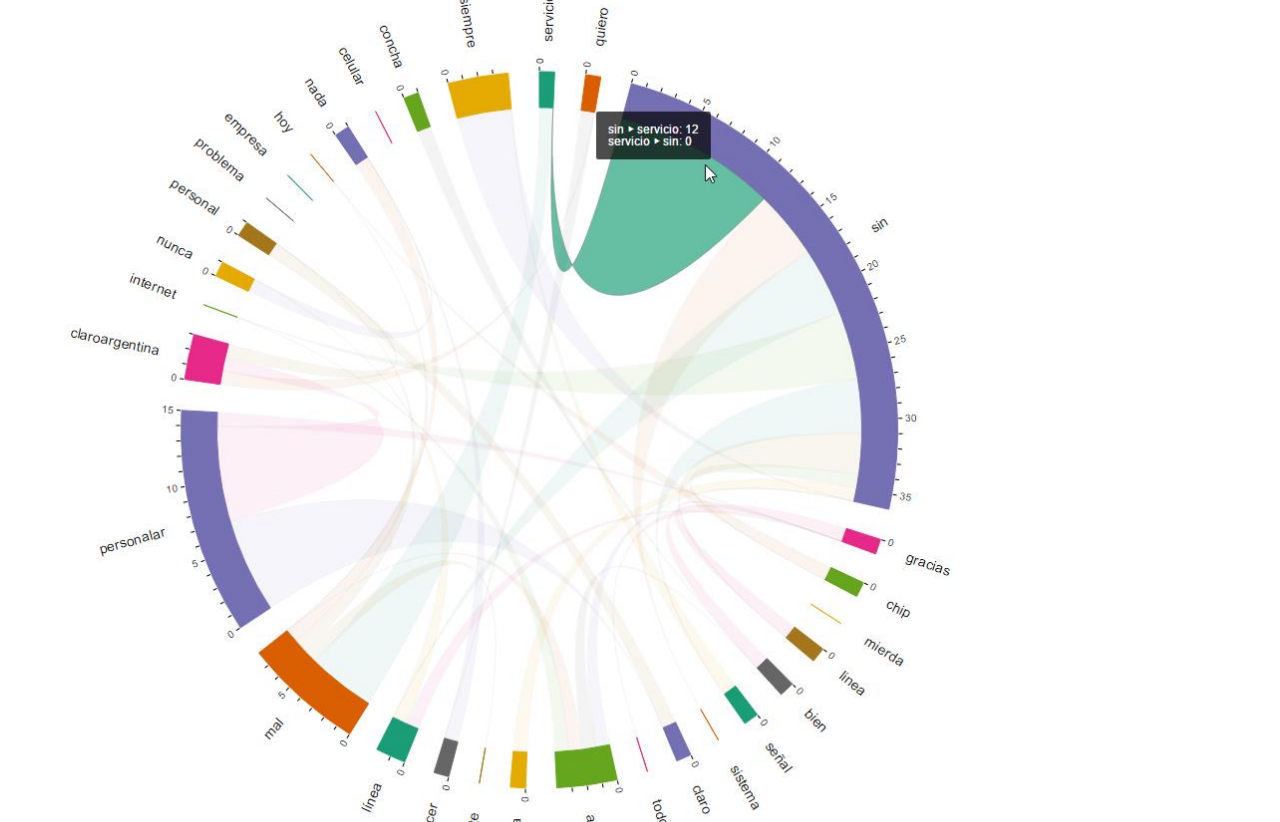
## EMPRESA TELCO ARGENTINA

- Plataforma Utilizada
  - HADOOP con Arquitectura Lambda. Procesamiento Batch y de Datos en Movimiento.
- Cluster
  - 48 Servidores 128/256Gb RAM, 12 Discos x 4/6Tb, 2 Procesadores-28 Cores c/u.
  - Expansión futura a 120 Servidores.
- Objetivos
  - Fidelización de Clientes
  - Análisis de la Competencia.
  - Predecir futuras bajas.
  - Sentimental Analysis de Redes Sociales y blogs.





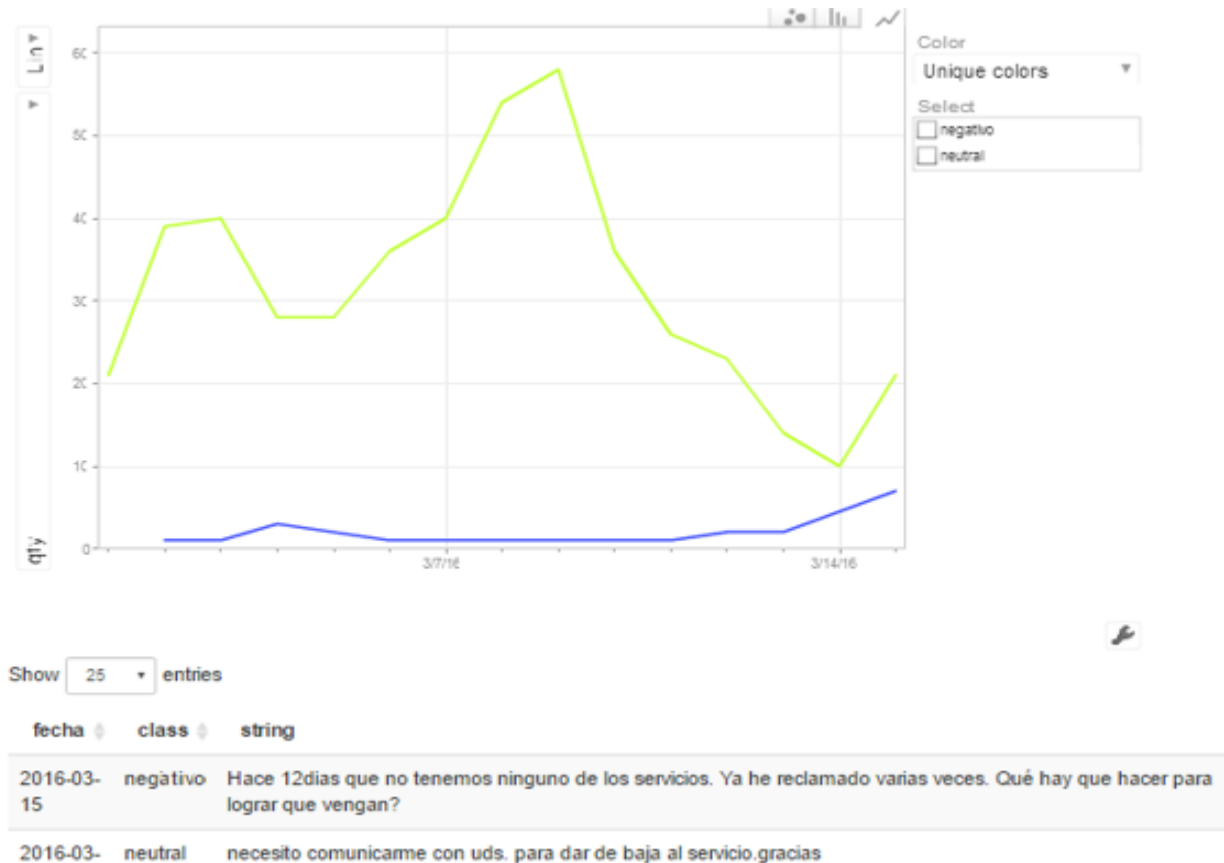
WordCloud  
Análisis de tweets real time.



## ChordGraph

# CASOS DE USO

## EMPRESA TELCO ARGENTINA



Sentimental Analysis  
Análisis de tweets real time.

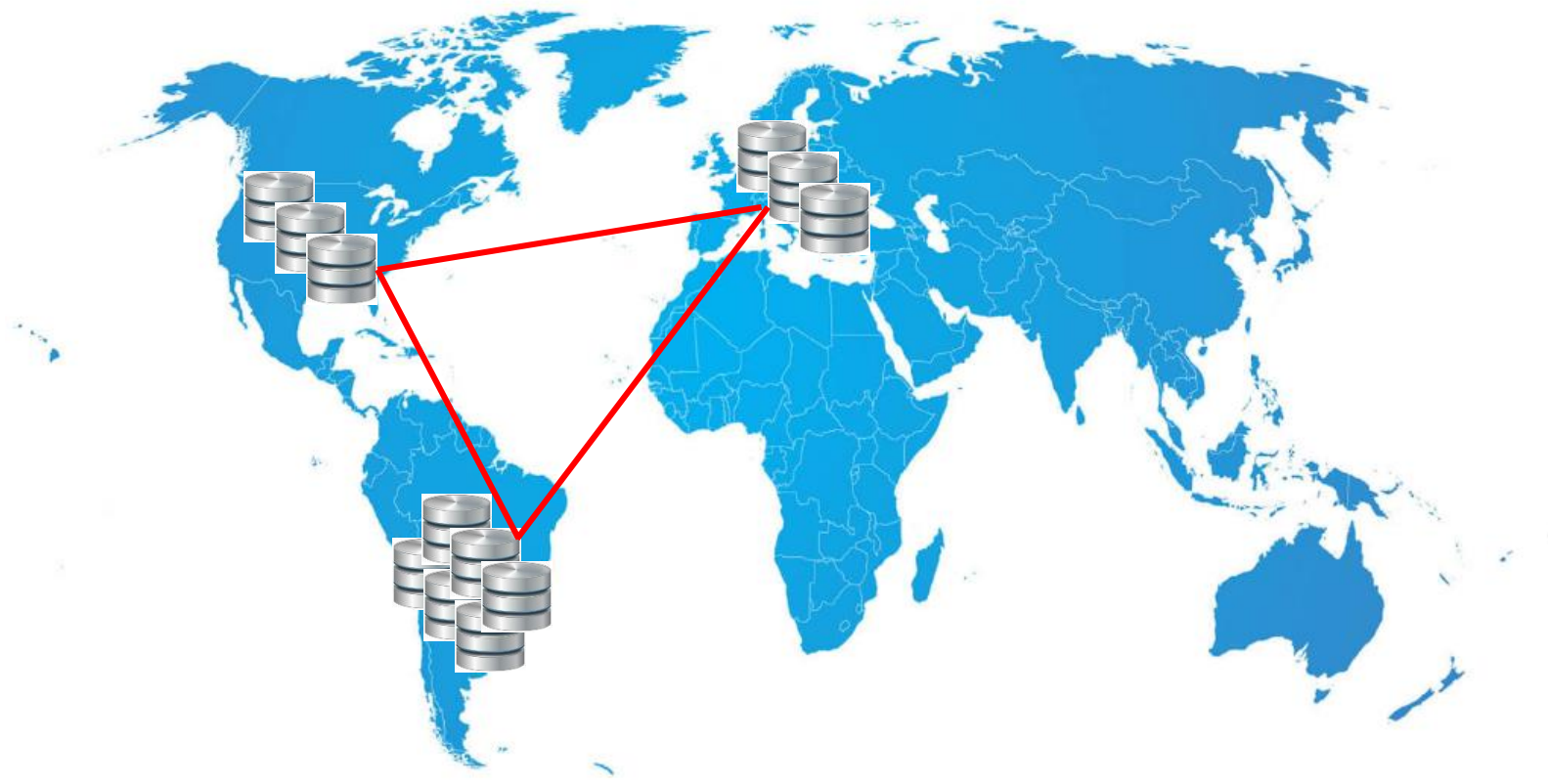
# CASOS DE USO

## EMPRESA FABRIL ARGENTINA

- Plataforma Utilizada
  - Base de Datos NoSQL MongoDB.
- Cluster
  - 3 Servidores 64Gb RAM, 8Tb Disco, 2 Procesadores XEON.
  - Expansión futura a 12 Servidores.
- Objetivos
  - Sistema de Recolección de sensores de máquinas de más de 20 empresas distribuidas en el mundo.
  - Aproximadamente 700 líneas productivas con 5 máquinas cada una.
  - Análisis y predicción de fallas.
  - Contar información Agregada global.

# CASOS DE USO

## EMPRESA FABRIL ARGENTINA



# CASOS DE USO

## GOBIERNO CENTROAMERICANO

- Plataforma Utilizada
  - HADOOP
  - Base de Datos NoSQL MongoDB.
  - Pentaho Suite
- Cluster
  - 12 Servidores 32Gb RAM, 6Tb Disco, 2 Procesadores XEON
  - Expansión futura TBD
- Objetivos
  - Mejorar la calidad y la eficiencia en las instituciones gubernamentales incluyendo percepción del ciudadano sobre la gestión de los servicios públicos
  - Analizar los artículos publicados en Blogs y Periódicos Digitales acerca de la gestión del gobierno.



# CASOS DE USO

Se identificaron cuatro grupos con impacto REAL en redes sociales:

## PRINCIPAL

## Incendio Forestal en sí.

## OPORTUNIDAD

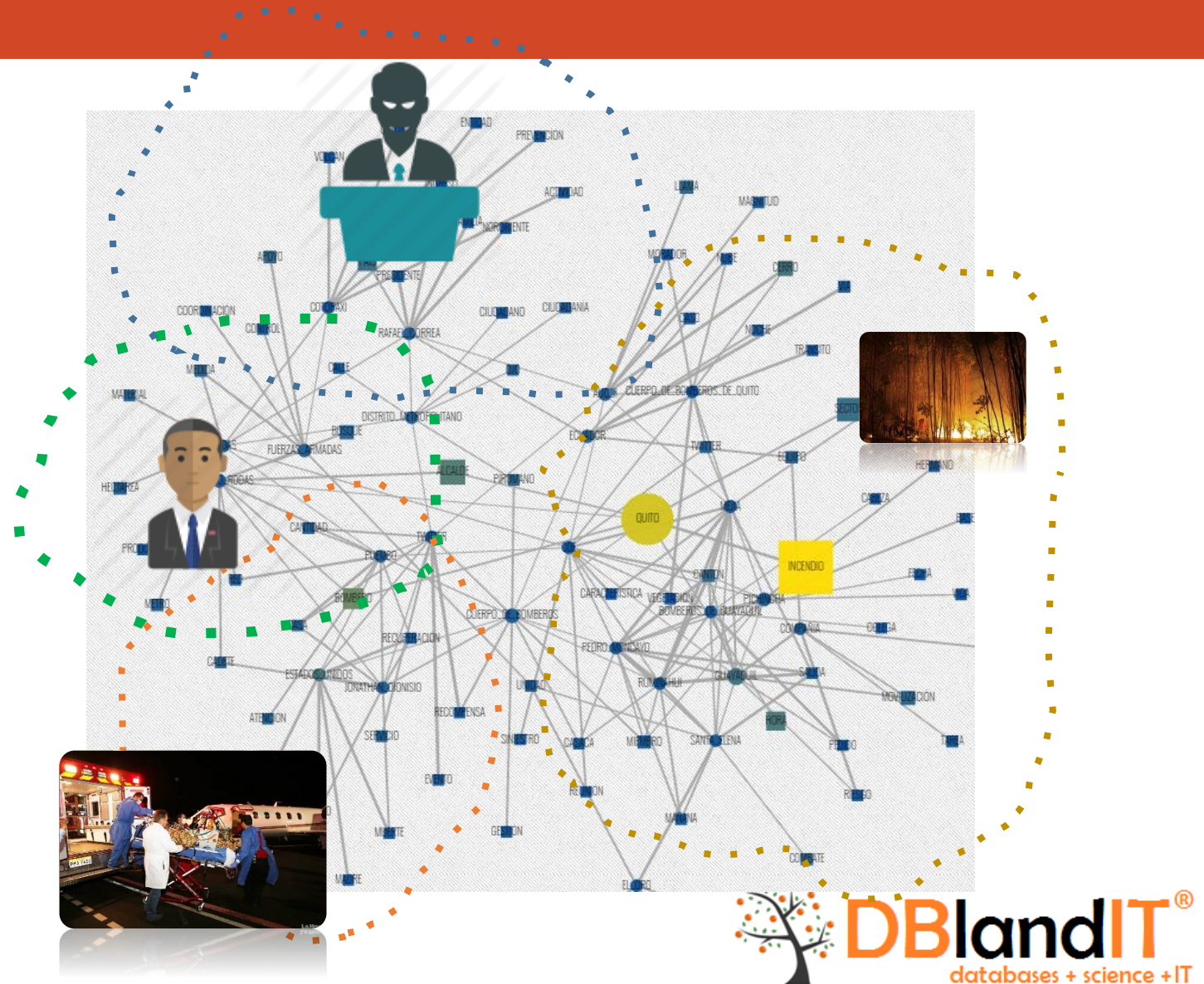
Bombero herido, reclamó traslado EE.UU.

## PERCEPCIÓN

Involucramiento del Intendente.

## PERCEPCIÓN

Involucramiento de Presidente.



# CASOS DE USO

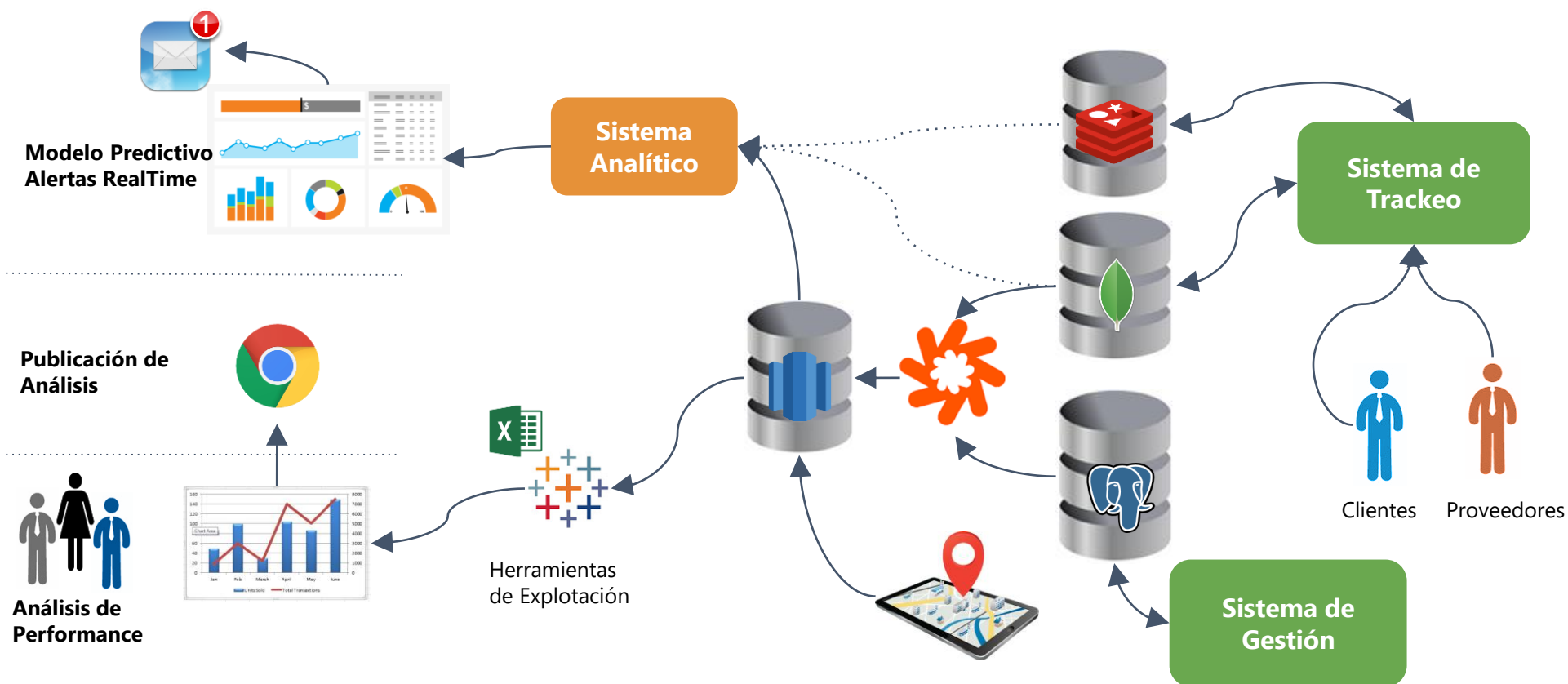
## EMPRESA ARGENTINA - MARKETING MOBILE

- Plataforma Utilizada
  - Base de Datos NoSQL MongoDB / Redis / Redshift.
  - Bases de Datos SQL PostgreSQL
  - Integración de Plataformas Data Virtuality
- Cluster
  - Hosting en AWS – Amazon Web Services
  - 8 Servidores 32Gb RAM, 500Gb Disco, 1 Procesador – 8/16Cores
  - Expansión futura TBD
- Objetivos
  - Sistema de Gestión y Trackeo 7 x 24.
  - Predicciones de Comportamiento
  - Performance de Campañas
  - Detección de Patrones de Fraude



# CASOS DE USO

## EMPRESA ARGENTINA - MARKETING MOBILE



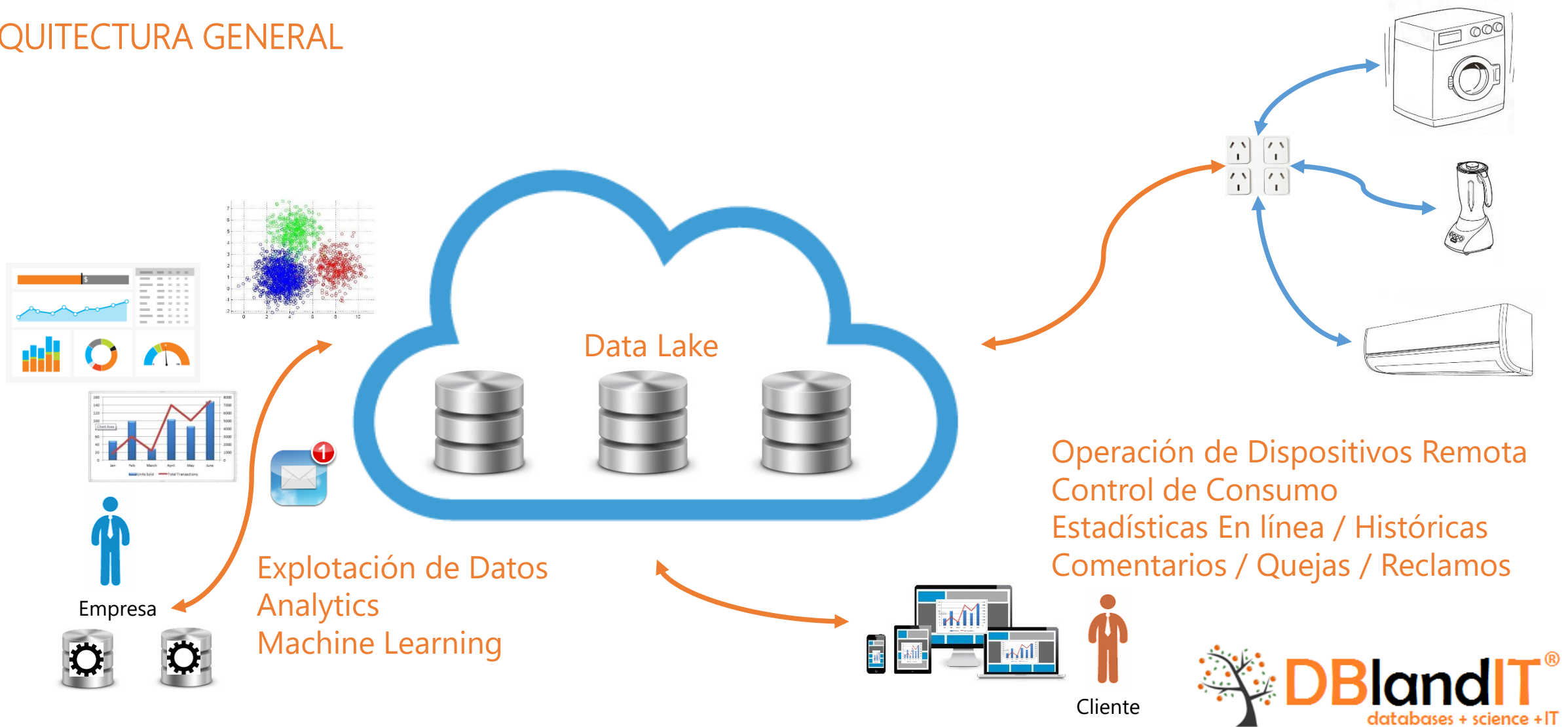
# SMART SHOE

## PROTOTIPO DESARROLLADO POR DBlandIT

- Desarrollo de un Tomacorriente Múltiple a 220v que puede ser operado de forma remota a través de Internet.
- Este dispositivo se conecta a una Plataforma de Big Data que almacena la información relacionada al uso de dicho dispositivo.
- Objetivos
  - Usuario Final:
    - Dar la posibilidad de operar de manera remota sus dispositivos conectados.
    - Contar con información en línea del consumo de los dispositivos y su consumo histórico, permitiéndole tomar decisiones que hagan a una reducción del consumo.
    - Poder agendar operaciones sobre los dispositivos que se ejecutarán a un determinado día y horario.
    - Informar al usuario de problemas existentes con el dispositivo, previendo futuras fallas.
  - Empresa:
    - Contar con información de cómo utilizan sus productos sus clientes.
    - Armar un perfil de sus clientes.
    - Predecir fallas y realizar servicios proactivos hacia sus clientes.
    - Realizar Mejoras de Producto en función del feedback recibido.

# SMART SHOE

## ARQUITECTURA GENERAL



Muchas Gracias !!!

**Ing. Juan Antonio Zaffaroni**  
**Founder & CEO de DBlandIT**  
**Docente de la Universidad Tecnológica Nacional**  
**[www.dblandit.com](http://www.dblandit.com)**  
**[jzaffaroni@dblandit.com](mailto:jzaffaroni@dblandit.com)**

