

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



**PROYECTO FIN DE CARRERA**

**Ingeniería de Telecomunicación**

"Estrategias, herramientas y métricas de posicionamiento de Big Data en laboratorios de investigación universitarios en redes sociales"

**Ismael Herrera De los Ríos**

**DICIEMBRE 2015**



"Estrategias, herramientas y métricas de posicionamiento de Big Data en laboratorios de investigación universitarios en redes sociales"

**AUTOR: Ismael Herrera De los Ríos**

**TUTOR: Carlos Minchola**

**Digital System Lab**

**Escuela Politécnica Superior**

**Universidad Autónoma de Madrid**

**DICIEMBRE de 2015**



## ***Agradecimientos***

***“El tiempo pone a cada uno en su lugar”.***

Posiblemente sean estas páginas las más difíciles de escribir dado que poco valor tiene lo que pueda decir frente al sacrificio que han hecho mis padres día tras día para que pueda llegar hasta aquí. Muchas han sido las horas, las alegrías, las decepciones pero también importante ha sido aprender que caer es el único modo de aprender a levantarse y saber valorar lo logrado.

Una vez un hombre sabio me dijo que las ingenierías las empiezan la inteligencia y las acaban la persistencia y hoy día esa frase cobra más sentido que nunca pero también debo remarcar que las ingenierías las haces tú y cada una de las personas que comparten tus días. Las horas de biblioteca acompañadas de bebidas energéticas, los meses estudiando que no siempre dan sus frutos son experiencias que sin el apoyo incondicional de mi familia y Paula no me habrían hecho hoy más fuerte.

Estos años no han sido solo un aprendizaje educativo, me han enseñado a trabajar en equipo y sobre todo a aprender que los proyectos no tienen horarios y que cuando no puedes más y tienes 6 horas para entregar una práctica que falla al completo es cuando nosotros los ingenieros damos lo mejor de nosotros. Creo que esos momentos de superación han sido los que han convertido a “compañeros” en amigos. Nuestros caminos profesionales nos han separado en el día a día pero hoy no puedo olvidar el apoyo de Raúl y Pili, esas dos personas que han sido la mejor influencia positiva y sobre todo un apoyo cuando más he necesitado consejo y también un ejemplo constante de trabajo y de que los apuntes se pueden hacer en algo más que en negro y azul. Hoy recuerdo también a Josito, una de esas personas que definen la lealtad, un compañero excepcional y la persona que me ha hecho falta cuando necesitaba que me dijese las cosas claras aunque no me gustase. Gracias también a Diego, mi casi vecino y una de esas personas cuyas alegrías las sientes como propias porque hoy te veo con orgullo en una carrera meteórica y parte de ese éxito lo siento como mío porque eres una de esas personas que aunque no vea todos los días sé que si algún día nos necesitamos no habrá más que descolgar el teléfono. Gracias a Tomás, fiel escudero en mi última etapa, los genios están locos y tú tienes mucha genialidad a ti te doy las gracias por las horas interminables hablando de política, de la vida y en

definitiva por romper juntos la monotonía con nuestras risas. Gracias a Agustín, porque empezamos haciendo prácticas de móviles y hoy tenemos el placer de llamarnos amigos, suerte siempre y no dudes de ti. Gracias Jorge por hacerme tener la “mirada del tigre”, por esas risas con Sergio, Guillermo y Álvaro y porque aunque os vea menos seguís siendo importantes.

Gracias a todos y cada uno de mis profesores, porque me habéis enseñado algo más que hacer transformadas de Fourier, me habéis exprimido, me habéis hecho confiar en mí y me habéis puesto el camino lo suficiente duro como para que hoy el premio tenga un mejor sabor. Gracias por la paciencia en las tutorías con preguntas infinitas, gracias por las clases los viernes que teníamos la cabeza en nuestros mundos y aun así dabais lo mejor de vosotros. Gracias.

Gracias a Eduardo Boemo, por confiar en mí en este proyecto, por presentarme a la electrónica digital uno de mis grandes amores y por hacer que las clases tengan ese trato especial y cercano. Gracias por esas tutorías en las que me enseñabas pero también te preocupabas por mí y me dabas consejos más allá de lo educativo. Gracias por compartir, enseñarme y darme el impulso suficiente para llegar aquí.

Gracias a Paula por darme la estabilidad que necesitaba, la calma en un mundo de locos y la comprensión y cariño cuando al mirarme en el espejo no me aguantaba ni yo mismo. Gracias por entender mis horas de biblioteca y mis charlas sobre tecnología. En definitiva gracias por aparecer y quedarte en mi vida.

Quiero también dar las gracias a Javier Oubiña y a toda la gente del Rectorado por darme la oportunidad de dar un granito de arena a la Universidad Autónoma de Madrid a la que tanto cariño tengo y por ser más que un jefe un líder que me ha dado la confianza necesaria para crecer en lo profesional y en lo personal. Gracias Javier por nuestras charlas, por los proyectos compartidos y sobre todo por la confianza y el valor que viste en mí.

Para finalizar quiero recordar a mis abuelos, una vez dijiste “mi nieto es ingeniero” y hoy aunque no de la misma manera sé que seguís aquí a mi lado. Hoy si sale todo bien os puedo decir que soy ingeniero y que igual que vosotros estabais orgullosos de mi cuando decidí emprender este camino yo estoy orgulloso de vosotros por haberos tenido a mi lado.

Porque si estás dispuesto a pasar a través de todas las luchas que tienes que pasar para llegar a donde quieres, ¿quién tiene el derecho de pararte?-(Rocky VI).

Humildemente a todos los que he conocido en el camino, los que están, se fueron o vendrán.

**Gracias.**

# INDICE DE CONTENIDOS

---

1. Introducción al Big Data .....	1
1.1 Introducción: Motivación.....	1
1.2 Introducción: Objetivos .....	2
1.3 Organización de la memoria .....	3
2. Estado del arte del Big Data .....	5
2.1 Big Data: Visión general.....	5
2.2 Big Data: Dimensiones de Big Data. ....	7
2.3 Big Data: Análisis de datos Estructurados y No estructurados .....	9
2.4 Big Data: ¿En qué punto estamos?.....	11
2.4.1 Big Data: La oportunidad de mercado en 2015 .....	12
2.5 Big Data: Analytics: el uso de Big data en el mundo real. ....	14
2.5.1 Big Data en los servicios financieros .....	14
2.5.2 Big Data en las compañías aseguradoras .....	18
2.5.3 Big Data en Ciencia: El científico de datos.....	19
2.5.4 Big Data en la medicina: personalizada, predictiva, preventiva y participativa". ..	20
2.5.5 Big Data en el deporte: El hombre contra la máquina.....	20
2.6 Big Data Nuevo paradigma de la Educación.....	22
2.6.2 Los cursos MOOC: España a la cabeza de Europa en creación de ellos. ....	22
2.6.3 Estrategias de innovación Educativa. ....	24



3. Diseño de una arquitectura Big Data .....	27
3.1 Fase 1: Recolección de datos. ....	27
3.2 FASE 2 Almacenamiento: Hadoop Distributed File System (HDFS).....	27
3.2.1 BASES DE DATOS RELACIONALES; SQL-NoSQL.....	29
3.3 Análisis de datos: MAPREDUCE.....	33
3.3 interfaz de usuario: VISUALIZACIÓN .....	36
3.3.1 Infografías con Easel.ly:.....	36
3.3.2 Infografías con Piktochart.....	38
3.3.3 Google Fusion Tables: .....	40
4 Sistema HADOOP .....	41
4.1 HADOOP DISTRIBUTED FILE SYSTEM (HDFS) 1.0.....	42
4.1.2. MAPREDUCE 1.0 .....	44
4.3. HADOOP 2.0 .....	49
4.3.1. HDFS 2.0: QUORUM JOURNAL MANAGER .....	49
4.3.2 MAPREDUCE 2.0 ARQUITECTURA YARN .....	50
5. Desarrollo Hadoop en Redes Sociales .....	53
5.1 Adquisición de datos y grabación: Flume .....	53
5.2 Extracción e integración de la información: AVRO .....	55
5.4 CONSULTAS DE DATOS: HIVE.....	61
Funcionamiento .....	62
5.6 Interpretación de datos.: Mahout.....	63
6. Big Data en las redes sociales de la UAM .....	69
6.1 ¿Qué relación tienen el Big Data y las redes sociales?.....	69
6.2 big data redes sociales: TWITTER UAM .....	72

6.2.1 twitter uam: análisis mediante socialbro.....	74
6.3 big data y redes sociales: Youtube y los MOOCs .....	82
6.3.1 Youtube Analytics. ....	84
6.4 big data y redes sociales: Facebook.....	87
6.5 Big Data: Analítica Web: Google Analytics y cookies.....	95
6.5.1 Analítica Web: SEO y SEM. ....	98
Proceso de optimización para buscadores SEO: .....	100
6.5.2 Analítica Web: Woorank valoración onpage.....	101
6.5.3 SEO para apps: ASO (App Store Optimization).....	109
6.5.4 Uso de blogs universitarios: Edublogs .....	112
<b>Conclusiones y trabajo futuro .....</b>	<b>115</b>
<b>Referencias .....</b>	<b>CXVII</b>
<b>GLOSARIO Big Data .....</b>	<b>CXX</b>
<b>Acrónimos del Marketing Digital.....</b>	<b>125</b>
<b>Hadoop: instalación y configuración .....</b>	<b>127</b>
<b>Formación y certificados .....</b>	<b>CXXXIII</b>
<b>Presupuesto solución Big Data redes sociales UAM .....</b>	<b>I</b>
<b>PLIEGO DE CONDICIONES .....</b>	<b>- 1 -</b>
<b>Condiciones particulares .....</b>	<b>- 6 -</b>

La presentación del proyecto se puede seguir en:

<http://ismaelherrerarios1.wix.com/bigdata>



El site sirve como portal para difundir los contenidos teóricos aprendidos en el presente PFC y exponer las líneas generales del proyecto englobadas en:

Big Data Analytics /Arquitectura Hadoop / Big Data aplicado a Redes Sociales/ Analítica Web

El site es a su vez un punto de encuentro de la diversa información del DSLab incluida alguna de las nuevas herramientas desarrolladas en el PFC como un espacio para la divulgación de aplicaciones móviles y difusión de noticias de eventos relacionados con el DSLab.

En la pestaña Presentación PFC se ubica el contenido de un Prezi con el contenido a mostrar el día de la presentación del proyecto recogiendo el histórico de la evolución del proyecto, los resultados obtenidos y las líneas de futuro en un recorrido visual y ameno que permita al lector obtener una visión global del trabajo realizado.

# INDICE DE FIGURAS

---

Figura 1 fuente IBM: Analytics: el uso de Big data en el mundo real

Figura 2: Escenarios reales de almacenamiento de datos en Big Data.

Figura 3: Tim McGuire director de McKinsey en Making data analytics

Figura 4 IBM Institute for Business Value

Figura 5: Global mobile payment spending from 2010 to 2017.

Figura 6: Paypal on Digital Wallet Market

Figura 7: África and Asia Mobile Payments.

Figura 8: Distribución de los MOOCs fuente “Comisión European MOOC Scoreboard”

Figura 9: Explicación Punto Neutro fuente Ministerio de Educación.

Figura 10: Arquitectura HDFs fuente “An introduction to the Hadoop Distributed File System” IBM

Figura 11: Ejemplo de Base de datos documentales fuente “IBM Developers Work”

Figura 12: Ejemplo modelo Grafos fuente “Fundación Telefónica”.

Figura 13: Ejemplo de algoritmo MapReduce fuente “IBM Developers Work Hadoop”

Figura 14: Interfaz de usuario Easelly.

Figura 15: Barra de herramientas de Easelly.

Figura 16. Ejemplo de infografía obtenido con Easel.ly sobre consumo de carburante.

Figura 17: Piktochart interfaz de usuario.

Figura 18: Piktochart herramientas de edición.

Figura 19: Piktochart ejemplo de infografía sobre Social Media.

Figura 20: Ejemplo de Fusion Tables de Google con tipo de mariposa fuente Google.

Figura 21: Ejemplo Hadoop HDFS fuente IBM Analytics What is

Figura 22: Arquitectura Map and Reduce fuente IBM Analytics What is Hadoop?

Figura 23: Workflow de la ejecución de un trabajo MapReduce en una arquitectura Hadoop.

Figura 24: Workflow de la ejecución de un trabajo MapReduce en una arquitectura Hadoop.

Figura 25: Evolución Hadoop 1.0 a Hadoop 2.0 fuente Big Bang Data Fundación Telefónica.

Figura 26: Hadoop 2.0 Distributed Data Storage fuente Big Bang Data Fundación Telefónica.

Figura 27: Arquitectura YARN fuente Big Bang Data Fundación Telefónica.

Figura 28: Arquitectura Flume que descarga información de las redes sociales Twitter además de leer ficheros de log.

Figura 29: Integración de Avro en la arquitectura anterior.

Figura 30: Ejemplo de súper columna en Cassandra.

Figura 31: Ejemplo de Keyspace fuente IBM Analytics.

Figura 32: Arquitectura Hive fuente Hortonworks.

Figura 33: Audiciente de Internet y número de usuarios en RRSS de 2014

Figura 34: Dispositivos de acceso a las diferentes redes sociales fuente Viko.

Figura 35: Monográfico adolescentes en redes sociales, fuente IAB.

Figura 36: análisis sociodemográfico y de uso de la red social Twitter fuente IAB.

Figura 37: Estadísticas por sexo e intervalos de edad de la red social Twitter fuente IAB.

Figura 38: AÑADIR

Figura 39: Interfaz de usuario de SocialBro.

Figura 40: analisis sociodemografico y de uso de la red social Youtube fuente IAB.

Figura 41: Top 10 de las Universidades españolas en Facebook.

Figura 42: Cabecera de la página institucional de Facebook de la UAM.

Figura 43: Ranking page de Universidades españolas en Facebook por LikeRank.

Figura 44: Ranking page de universidades bajo estudio en 2015 internacionales y privadas.

Figura 45: Análisis de los datos obtenidos de una publicación en Facebook.

Figura 46: Análisis de interacción de la comunidad en Facebook.

Figura 47: Infografía por sexo y edad de la comunidad de usuarios UAM en Facebook.

Figura 48: Analisis geografico de los usuarios de Facebook.

Figura 49: Imagen del posicionamiento SEO de una página web.

Figura 50: Imagen del posicionamiento SEM de una página web.

Figura 51: Report de Google Trends para Universidad Autónoma de Madrid.

Figura 52: Report de Google Trends con consultas frecuentes para Universidad Autónoma de Madrid.

Figura 53: Report de Google Trends con geolocalización del interés por “Universidad Autónoma de Madrid”.

Figura 54: Mapa geográfico del tráfico web de la página de la Universidad Autónoma de Madrid.

Figura 55: App Comb Circuits DSLab

Figura 56: Descripción App Comb Circuits DSLab

Figura 57: Categoría y tipo de app Comb Circuits DSLab

Figura 58: Capturas de la app Comb Circuits DSLab

Figura 59: Valoración de la app Comb Circuits DSLab

Figura 60: Landing page de Digital System Laboratory.

Figura 61: Estadísticas a 7 de noviembre de tráfico y posicionamiento.

# INDICE DE TABLAS

---

Tabla 1: Tabla “Información cliente móvil”.

Tabla 2: “Base de Datos NoSQL Clave Valor cliente.”

Tabla 3: Ejemplo consulta SQL

Tabla 4: Ejemplo sistema de identificación de usuarios. Horizontal.

Tabla 5: Ejemplo sistema de identificación de usuarios. Vertical.

Tabla 6: Ejemplo MapReduce, fichero inicial con fecha, nombre y puntuación de usuarios.

Tabla 7: Ejemplo MapReduce: Archivo Map modelo clave-valor.

Tabla 8: Ejemplo Mapreduce: Resultado Map key/value

Tabla 9: Ejemplo MapReduce, proceso shuffle and sort.

Tabla 10: Ejemplo MapReduce, resultado proceso reduce.

Tabla 11: Ejemplo de SupercolumnFamily.

Tabla 12: Mahout frente a métodos tradicionales por usuario.

Tabla 13: Comparativa de datos obtenidos Twitter Enero 2015vs Diciembre 2014 vía Socialbro.

Tabla 14: Red score evolution durante el mes de Enero 2015.

Tabla 15: Ratios de engagement de Twitter de la Universidad Autónoma de Madrid.

Tabla 16: Evolución de seguidores en Enero 2015 y previsión Febrero de Twitter de la Universidad Autónoma de Madrid.

Tabla 17: Grafico detallado de la interacción con los usuarios con nuestras publicaciones.

Tabla 18: Mapa de palabras más utilizadas por la comunidad de seguidores de la UAM.

Tabla 19: Tagcloud de las ubicaciones de la comunidad de usuarios de la Universidad Autónoma de Madrid.

Tabla 20: Estadísticas de los idiomas de nuestra comunidad y las zonas horarias de mayor actividad de la UAM.

Tabla 21: Estadísticas de visualizaciones de los videos de Enero 2015.

Tabla 22: Análisis sociodemográfico de la comunidad de usuarios de Youtube de la UAM.

Tabla 23: Comportamiento del canal durante el mes de enero de 2015.

Tabla 24: Resumen general de la página de Facebook por usuarios e interacción.

Tabla 25: Page Performance y Post by Pages obtenidos con LikeALyzer.

# 1. INTRODUCCIÓN AL BIG DATA

---

Memoria del **Proyecto Fin de Carrera** de los estudios de Ingeniería Superior en Telecomunicación, impartidos por la Escuela Politécnica Superior de la *Universidad Autónoma de Madrid*, del alumno Ismael Herrera De los Ríos.

## 1.1 Introducción: Motivación

Los teléfonos móviles y el desarrollo de nuevas aplicaciones de electrónica digital generan cantidades enormes de datos. Los analistas del sector de las Tecnologías de la información y la Comunicaciones consideran por ello al fenómeno Big data la nueva frontera de la innovación, la competencia y la productividad **“McKinsey, 2011”**. Algunos datos que lo corroboran:

- El **Big Data** generará 4,4 millones de empleos en 2015 en todo el mundo (*Gartner*).
- El 90% de la información generada a nivel mundial se ha producido en los últimos 2 años. (*IAB 2014*).
- El PIB de Europa crecerá 1.9 % adicional gracias al *Big Data*. (**IBM**)

La inversión en *Big Data* será de 122 millones de euros en 2015.

Desde el punto de vista del estudiante, la realización de este PFC es interesante a nivel formativo, pues la mayoría de las tecnologías estudiadas son tan recientes que apenas tienen cabida en los planes de estudios vigentes. Además, en el plano empresarial las compañías están empezando a adoptar soluciones *Big Data* dentro de sus infraestructuras, por lo que es un punto clave dentro de la evolución y crecimiento de éstas.

*Referencias McKinsey, “Big data: The next frontier for innovation, competition, and productivity”*

*Referencias IAB. “Big Data Empowerment: Promoting Civic Engagement”*



## 1.2 Introducción: Objetivos

La tecnificación del análisis de datos ha llevado a la implementación y desarrollo de varias aplicaciones y herramientas que permiten el tratamiento y evaluación de los mismos. Saber medir todo lo que hacemos en plazos de tiempo mínimos nos permite tomar decisiones en base a unos KPIs (Indicadores clave de rendimiento objetivable) permitiendo una evaluación de resultados y toma de decisiones más dinámica.

En el presente PFC se va a utilizar todas estas herramientas para realizar la monitorización de los datos que circulan en las redes sociales corporativas de la **Universidad Autónoma de Madrid** y en particular, del *Digital System Lab*. El objeto es conocer la utilidad de los datos obtenidos mediante sistemas *Big Data* en redes sociales y emprender acciones estratégicas para que los estudiantes tengan mayor conocimiento y acceso a los materiales y canales ofrecidos para su aprendizaje.

- **Estudio teórico**

Análisis de las características, herramientas, tecnologías y estándares relacionados con *Big Data*. Sectores que implementan en la actualidad soluciones *Big Data* en sus estrategias tales como el sector bancario, el deporte de elite o la educación universitaria.

Análisis de la tecnología *Big Data*: En este apartado se define un marco general de la aparición y definición de la tecnología *Big Data*. Para ilustrar mejor el funcionamiento de los sistemas se exponen implementaciones reales de estos sistemas.

Arquitectura hardware de un sistema *Big Data*: En este apartado se desarrolla las diferentes fases de un sistema *Big Data* con las tecnologías actuales especificando los criterios de selección teniendo como objetivo conocer su implementación en los distintos escenarios universitarios.

- **Estudio técnico**

En este apartado se analizan distintas implementaciones reales de analítica de *Big Data*. Se realiza una descripción de las más importantes herramientas para la captura y análisis de grandes volúmenes y se realiza el análisis de los datos que circulan por las redes sociales de la Universidad Autónoma de Madrid. En este apartado es de vital importancia constatar mi colaboración profesional con el Delegado del Rector para Imagen e identidad *Don Javier Oubiña* gracias al cual he podido acceder a un escenario real de análisis de datos. Las certificaciones que poseo por parte de Google han sido determinantes a la hora de escoger entre diferentes tecnologías.

## **1.3 Organización de la memoria**

<b>1. Introducción al Big Data .....</b>	<b>1</b>
<b>2. Estado del arte del Big Data .....</b>	<b>5</b>
<b>3. Diseño de una arquitectura Big Data.....</b>	<b>27</b>
<b>4 Sistema HADOOP.....</b>	<b>41</b>
<b>5. Desarrollo Hadoop en Redes Sociales.....</b>	<b>53</b>
<b>6. Big Data en las redes sociales de la UAM .....</b>	<b>69</b>
<b>Conclusiones y trabajo futuro .....</b>	<b>115</b>
<b>Presupuesto solución Big Data redes sociales UAM .....</b>	<b>I</b>



## 2. ESTADO DEL ARTE DEL BIG DATA

---

### 2.1 Big Data: Visión general

**¿Qué es *Big Data*?** en términos generales podría ser la tendencia tecnológica hacia un nuevo enfoque en el análisis y toma de decisiones, utilizada para describir enormes cantidades de datos que tomaría demasiado tiempo y sería muy costoso usando bases de datos relacionales para su arquitectura.

Según [1], el término aplica a la información que no puede ser procesada o analizada mediante procesos tradicionales. Para [2], *Big Data* son “cantidades masivas de datos que se acumulan con el tiempo que son difíciles de analizar y manejar utilizando herramientas comunes de gestión de bases de datos”, y para [3], *Big Data* se refiere “al tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales”.

En “*Forrester*” se define *Big Data* como las técnicas y tecnologías que hacen que sea económico hacer frente a los datos a una escala extrema. *Big Data* trata de tres cosas:

- 1) Las técnicas y la tecnología, lo que significa que la empresa tenga personal, el cual tenga gran representación y análisis de datos para tener un valor agregado con información que no ha sido manejada.
- 2) Escala extrema de datos que supera a la tecnología actual debido a su volumen, velocidad y variedad.
- 3) El valor económico, haciendo que las soluciones sean asequibles y ayuden a la inversión de los negocios”.

**¿Cuándo nace *Big Data*?** En plena burbuja de Internet donde el número de páginas Web crecía exponencialmente. Los usuarios empezaban a acceder a Internet para encontrar información consultando páginas de índices como Yahoo y motores de búsquedas basados en palabras clave como AltaVista. En ocasiones los resultados de

las búsquedas realizadas no resultaban relevantes y manifestaban problemas de rendimiento en situaciones de análisis de muchas páginas.

Como solución a ese problema de búsquedas en 1996 *Sergey Brin y Lawrence Page* comienzan la creación de un motor de búsqueda llamado **BackRub en la Universidad de Stanford**. Escriben el artículo "***The Anatomy of a Large-Scale Hypertextual Web Search Engine***" donde presentan un prototipo de un motor de búsqueda, llamado *Google*. *Google* presentaba un nuevo modelo de análisis de datos en búsquedas utilizando la estructura presente en hipertextos (documentos que hacen referencias a otros documentos mediante hiper-enlaces). En dicho artículo los creadores de *Google* muestran su preocupación por la escalabilidad del motor de búsqueda y las limitaciones que tienen las arquitecturas centralizadas de índices para dar unos tiempos de respuesta aceptables.

Como respuesta a este problema de escalabilidad y rendimiento, los investigadores de *Google* desarrollan el sistema de ficheros distribuido ***Google File System*** (2003) y el framework de procesamiento de datos distribuido *MapReduce* (2004), que permiten el análisis de enormes volúmenes de conjuntos de datos. *Entramos en la Era del Big Data*.

**¿Cuál es la tecnología en Big Data?** En 2005, Yahoo crea un proyecto open source llamado *Apache Hadoop* que se convirtió en un estándar para las arquitecturas de *Big Data*.

Todos los grandes fabricantes de software (IBM, Oracle, Microsoft, etc.) incorporan a sus productos diversos componentes de *Apache Hadoop* para poder vender sus productos como *Big Data*.

Existen otras arquitecturas consideradas *Big Data* como por ejemplo *Google BigQuery* y *Amazon DynamoDB* que tienen una aproximación distinta a *Apache Hadoop*, aunque empiezan a tener funcionalidades parecidas.

### **¿Cuánto valen nuestros datos personales?**

Según un estudio de *Mediabrand Marketing Sciences*, los usuarios están dispuestos a vender su información personal por unos 694 euros al año. Un 1% de todos los encuestados considera que las empresas tienen capacidad para cuidar sus datos, 6 de

cada 10 temen que los datos personales que pueden tener las empresas acaben en manos inadecuadas.

El 60% asegura que se sienten "extraños" ante la idea que las marcas los estén persiguiendo por la red para quedarse con su información y sus pautas de comportamiento. Y, puestos a ser seguidos, mejor vender los datos: 3 de cada 10 estaría más que dispuesto a hacerlo. Y aunque no se trate de venderlos, los consumidores parecen más dispuestos a compartir información personal si hay algún tipo de beneficio asociado o publicidad personalizada.

El estudio *The value of our digital identity*, elaborado por The Boston Consulting Group, señalaba que los datos personales de los consumidores tendrían un efecto muy importante en la economía europea: el uso de esta información reportaría un beneficio, según las estimaciones, de 1 billón de euros en 2020 a la economía de la Europa de los 27.

En España, los consumidores consideran que el nombre completo y la fecha de nacimiento tienen un valor de 14,25 euros, el número de móvil de 16,65 euros, la geolocalización de 15,87 euros, los ingresos anuales de 16,54 euros y el estado civil de 12,72 euros. De media, los datos personales les saldrían a las empresas por 15,27 euros, que subirían hasta los 17,90 en el caso de aquellas compañías con las que los consumidores no están familiarizados.

## 2.2 Big Data: Dimensiones de Big Data.

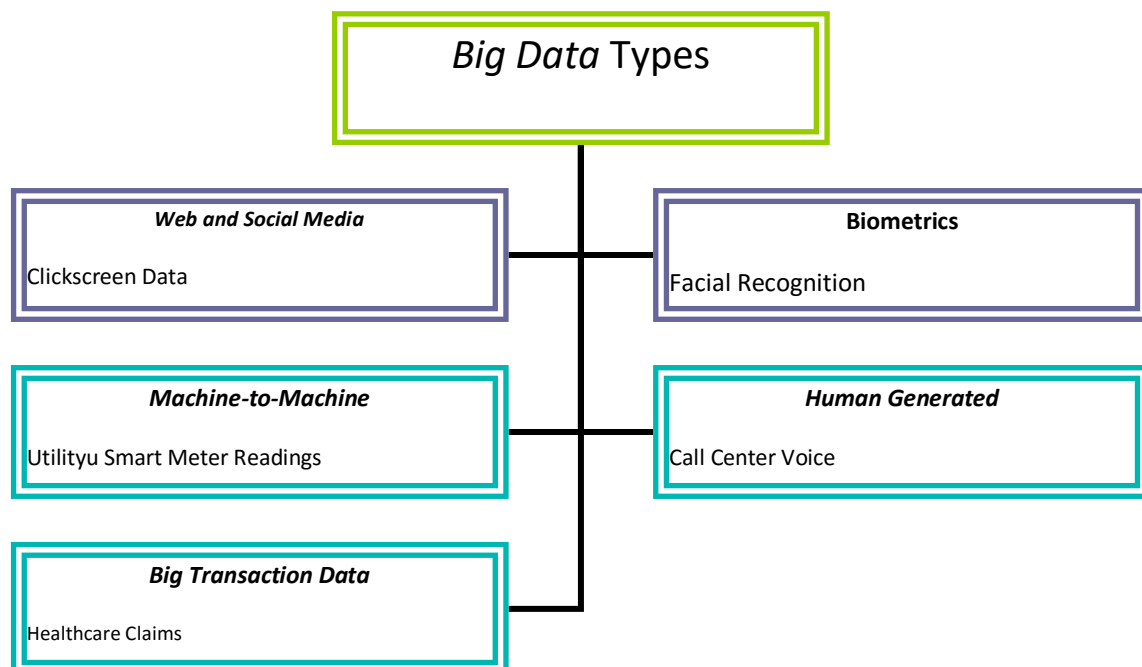
**Volumen:** El volumen de los datos almacenados en los depósitos de las empresas ha pasado de ocupar megabytes y gigabytes a "petabytes". En el año 2000 se generaron 800.000 petabytes (PB) y se espera que en 2020 esta cifra alcance los 35 zettabytes (ZB).

"En 2016 habrá 18.9 billones de dispositivos conectados, esto conllevaría que el tráfico de datos móviles alcance 10.8 Exabytes mensuales o 130 Exabytes anuales. Este volumen de tráfico equivale a 33 billones de DVDs anuales." *Naciones unidas*

Las redes sociales son una de las mayores fuentes de datos, *Twitter* genera 12 terabytes (TB) diariamente, y de Facebook, 10 TB de datos diarios.

**Valor:** Ser capaces de tener el criterio adecuado para analizar solo la información útil en tiempos mínimos y a un coste aceptable.

**Variedad:** La variedad de datos ha explotado, pasando de ser datos almacenados y estructurados, guardados en bancos de datos empresariales, a ser desestructurados, semiestructurados, audio, video, XML, etc



**Velocidad:** La velocidad refleja la frecuencia con la que los datos se generan, almacenan y comparten.

La velocidad del proceso y captura de datos ha aumentado significativamente. Los modelos basados en inteligencia de negocios generalmente suelen tardar días en procesarse, frente a las necesidades analíticas "casi" en tiempo real de las soluciones financieras.

## Dimensiones de big data

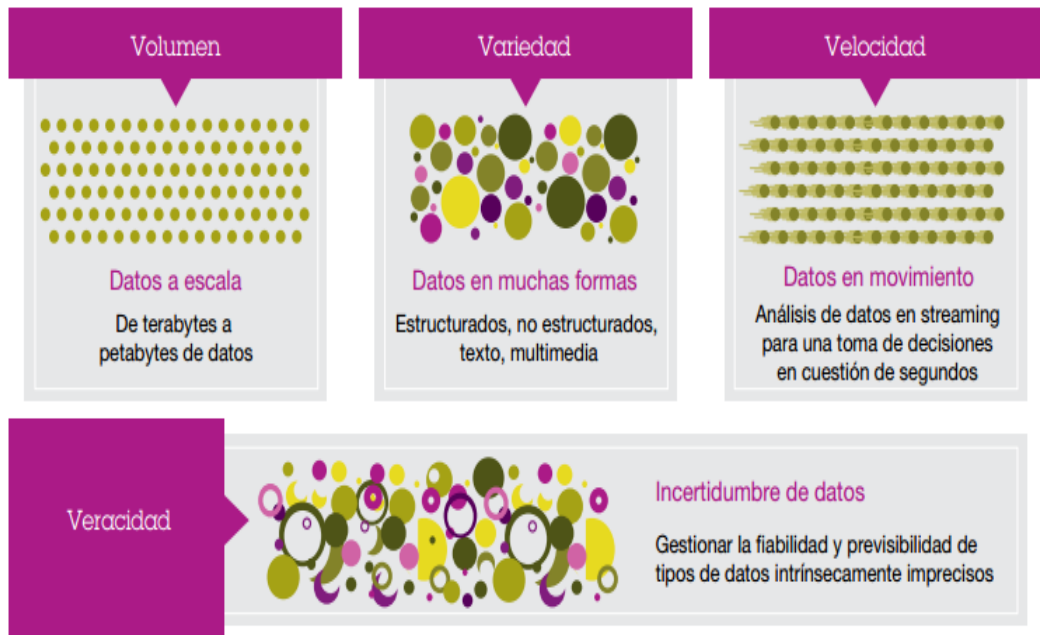


Figura 1 fuente IBM: Analytics: el uso de Big data en el mundo real

## 2.3 Big Data: Análisis de datos Estructurados y No estructurados

**Datos estructurados:** *Big Data* no solo es válido para el análisis de datos no estructurados, desde su concepción se consideró para almacenar y analizar datos estructurados. *Google* fue el pionero en implementar una tecnología para ese tipo de datos llamada Bigtable. Bigtable se utiliza para guardar datos de varios proyectos en *Google*, como por ejemplo los índices de búsqueda *Google Earth* y *Google Finance*.

### Ejemplo de información estructurada

Nombre	Apellidos	Email	Compañía móvil
Ismael	Herrera De los Ríos	<a href="mailto:Ismael.herrera.rios@gmail.com">Ismael.herrera.rios@gmail.com</a>	Vodafone

Tabla 1: Tabla "Información cliente móvil".

**Datos semi-estructurados:** Datos que no se limitan a campos determinados, pero que contiene marcadores para separar los diferentes elementos. Es una información



poco regular como para ser gestionada de una forma estándar. Estos datos poseen sus propios metadatos semiestructurados que describen los objetos y las relaciones entre ellos, y pueden acabar siendo aceptados por convención. Un ejemplo es el HTML, el XML o el JSON.

**Datos no estructurados:** “Datos no-estructurada” son aquellos datos que no están organizados bajo el *Modelo de Datos Relacional* definido por Edgar Codd en 1970. Algunos ejemplos de datos no estructurados son los archivos de texto, documentos (PDF, Word), imágenes, audio y video.

La tecnológica actual permite el análisis de datos estructurados o semi-estructurados pero los usuarios cada vez generan más datos no estructurados por ello se busca evolucionar reduciendo los retardos en el manejo de dichos datos.

Existen múltiples beneficios en el análisis de *Big Data* para las organizaciones, demostrados en la encuesta realizada por **TDWI (The Data Warehousing Institute)**, cuando se preguntó: “¿Cuál de los siguientes beneficios se produciría si la organización implementa alguna forma de análisis de Big Data?”. El 61% respondió que influye de manera social; el 45%, que habrá más puntos de vista de negocio; el 37% se inclinó por las decisiones automatizadas en los procesos en tiempo real; el 29% mencionó que se mejoraría la planificación y la previsión, y el 27%, que se entendería el comportamiento del consumidor.

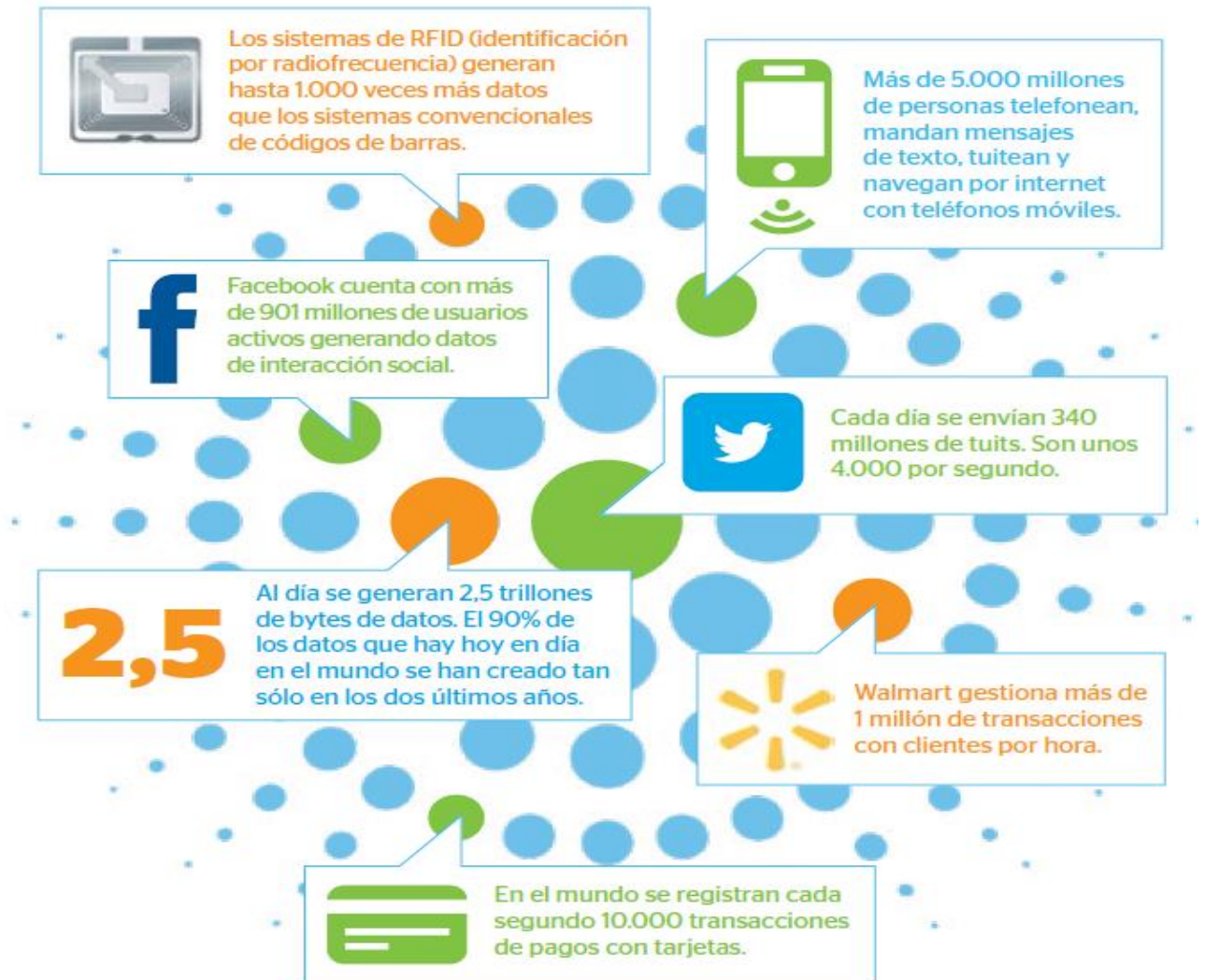
Base de Datos NoSQL Clave-Valor	
Clave	Valor
1	Nombre: Ismael, Nacionalidad: Española
2	Nombre: María, Nacionalidad: Española
3	Nombre: Pedro, Nacionalidad: Española

Tabla 2: “Base de Datos NoSQL Clave Valor cliente.”

El volumen de datos y la tasa de crecimiento de los datos no estructurados son muy superior al de los datos estructurados. De acuerdo con *Gartner*, la tasa anual de crecimiento de datos es del 40 a 60 por ciento. (Informe 2012). Dado el ingente volumen de datos, se impone el almacenamiento *Cloud*.

## 2.4 Big Data: ¿En qué punto estamos?

*Big Data* ya no es una promesa ni una tendencia. El sector empresarial se ha lanzado a la carrera de implementar *Big Data* en sus estrategias de negocio en todos sus proyectos y productos. El almacenamiento y análisis de grandes volúmenes de datos de diversas fuentes y a cada vez mayor velocidad se ha convertido en un elemento de diferenciador.



**Fuentes:**

SAS | Big Data: [www.sas.com/big-data](http://www.sas.com/big-data).  
IBM | Big Data at the Speed of Business  
McKinsey Global Institute | Big Data: The next frontier for innovation, competition, and productivity, junio 2011.

Figura 2: Escenarios reales de almacenamiento de datos en Big Data.

## 2.4.1 Big Data: La oportunidad de mercado en 2015

La encuesta anual sobre *Big Data* que realiza Gartner revela que el 73 % de los encuestados ha invertido o tiene planes para invertir en *Big Data* durante el año 2015.

La velocidad con la que crecen los datos se estima que se duplica cada dos años (Informe EMC. 2011 Digital Universe Study) y, según el informe Ericsson Mobility Report (agosto 2014), su tráfico creció un 60% entre el segundo trimestre de 2013 y el de 2014. El McKinsey Global Institute (MGI) en su informe *Big data: la próxima frontera de la innovación, la competencia y la productividad*, argumenta que los datos se están convirtiendo en un factor de producción, como lo es el capital físico o humano.

Un informe de Oracle advierte que probablemente en el sector retail, si se aplicaran técnicas *Big Data*, se podrían incrementar los márgenes en un 60% y los costes de fabricación y mantenimiento de las empresas se verían reducidos a casi la mitad.

Pero también se hace ***Big Data*** en medicina, en física... “En el CERN tienen el mayor sistema de *Big Data* del mundo: 300 Petabytes en 2013, lo que equivale a unas tres veces toda la información que, desde sus inicios, ha generado Facebook”, indica Pablo Casado, *chief architect* de Incubio, incubadora española especializada en startups de *Big Data*.

Según la presentación de Gartner “Top Technology Predictions for 2013 and Beyond”:

2. La demanda de gastos en servicios de *Big Data* se calcula que alcanzará los 132.300 millones de dólares en 2015.
3. En 2015 la demanda de *Big Data* supondrá 4,4 millones de puestos de trabajo en todo el mundo, pero sólo un tercio de ellos estarán cubiertos.

La demanda generara 550.000 empleos en servicios externos durante los próximos 3 años.

Otros 40.000 empleos los generarán vendedores de software durante los próximos 3 años.

El problema del almacenamiento en el siglo XX acabó, gracias a los avances como el Cloud computing. La nube es uno de las mayores revoluciones por la mejora en la seguridad y accesibilidad a los datos. Acceder en streaming a tiempo real a contenidos almacenados en la nube, está siendo muy utilizado en vídeos de presentaciones y conferencias.



“La mayor parte de los sectores reconocen ya que Big Data y el análisis de datos pueden disparar la productividad, hacer que los procesos sean más visibles y mejorar las predicciones sobre el comportamiento. Los análisis posteriores definirán la diferencia entre los perdedores y los vencedores en el futuro,”

dice Tim McGuire,  
director de McKinsey.



Ver video 

Fuente: BMckinsey | Making data analytics work: Three key challenges, marzo 2013.

Figura 3: Tim McGuire director de McKinsey en Making data analytics

## 2.5 Big Data: Analytics: el uso de Big data en el mundo real.

Cada vez son más las áreas que se benefician del *business intelligence*. Según un estudio de **Booz and Company** las áreas que han realizado implementaciones reales de *Big Data* son las siguientes:

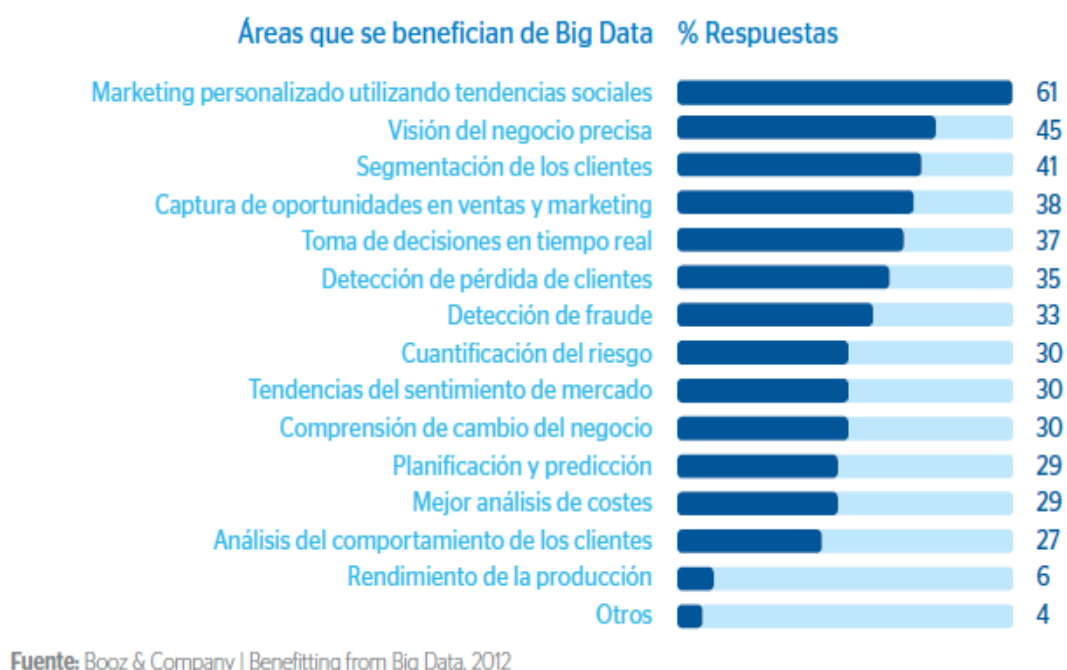


Figura 4 IBM Institute for Business Value

### 2.5.1 Big Data en los servicios financieros

*"Los bancos del futuro serán no solo bancos, sino grandes compañías de información".*

**Francisco González, Presidente de BBVA.**

Los datos de los servicios financieros expresan las decisiones de compra de los consumidores. Desde la banca tradicional hasta los grupos de inversores y compañías especializadas en créditos rápidos logran mediante *Big Data Analytics*, disponer de información actualizada y a tiempo real sobre el estado de los mercados y de los riesgos que se asumirán en cada operación que se platee.

Según **Retail Banking 2020, evolution or revolution?** El primer paso consiste en recobrar la confianza en el cliente. El informe plantea que la banca tanto retail como privada deben volcarse en conocer al cliente, conocer sus objetivos financieros y alinear sus productos financieros con ellos. Para ello es necesario dotar al cliente de una cultura financiera y para ello el primer avance es el acceso a la información.

FUTURO DE LAS ENTIDADES BANCARIAS

La tendencia inicial en las entidades financieras era aumentar el número de oficinas bancarias para tener conocimiento del cliente en primera persona. La tendencia actual es la supresión de las oficinas bancarias por la aparición de los smartphones y la viralidad de las redes sociales como medio de atención al cliente. Los ciudadanos tienen mayor información, mayor consciencia del riesgo y mejor cultura del gasto y del ahorro. Así, cada vez utilizamos más los dispositivos móviles para hacer muchas de las cosas que hacíamos en las oficinas bancarias y la tendencia en su uso:

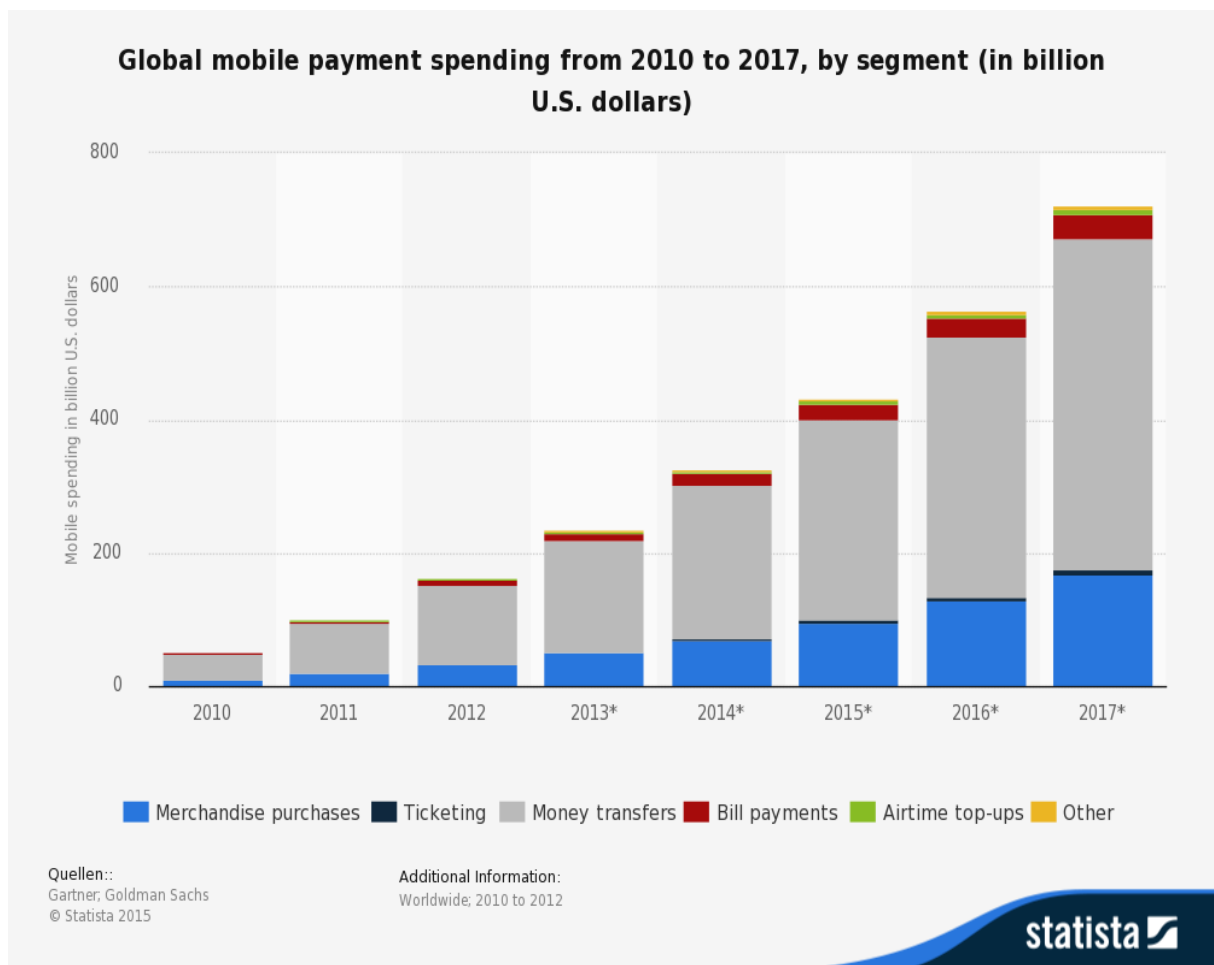


Figura 5: Global mobile payment spending from 2010 to 2017.

Francisco González, presidente del banco BBVA, señala a *Google*, *Amazon* y *Facebook* como los principales rivales de la banca y afirma que cuentan con herramientas con un gran potencial como *PayPal*. Recientemente, se ha añadido *Apple* con su **Apple Pay**.

El siguiente grafico demuestra que esa realidad ya existe en Estados Unidos donde PayPal es la aplicación más usada de pago.

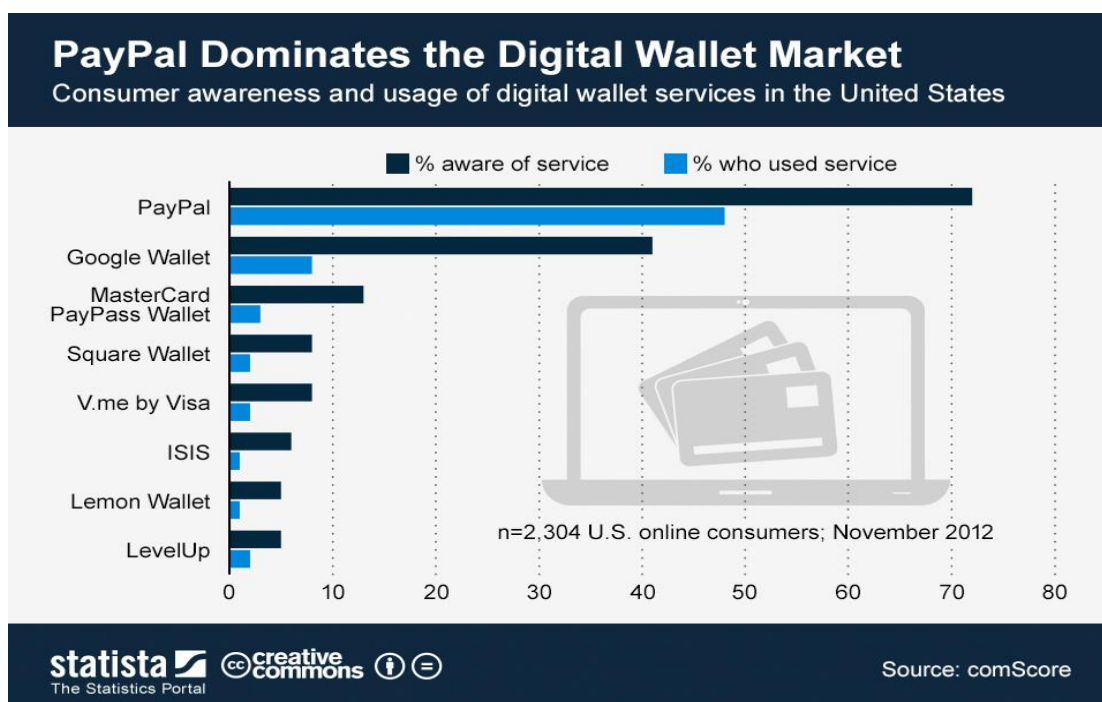
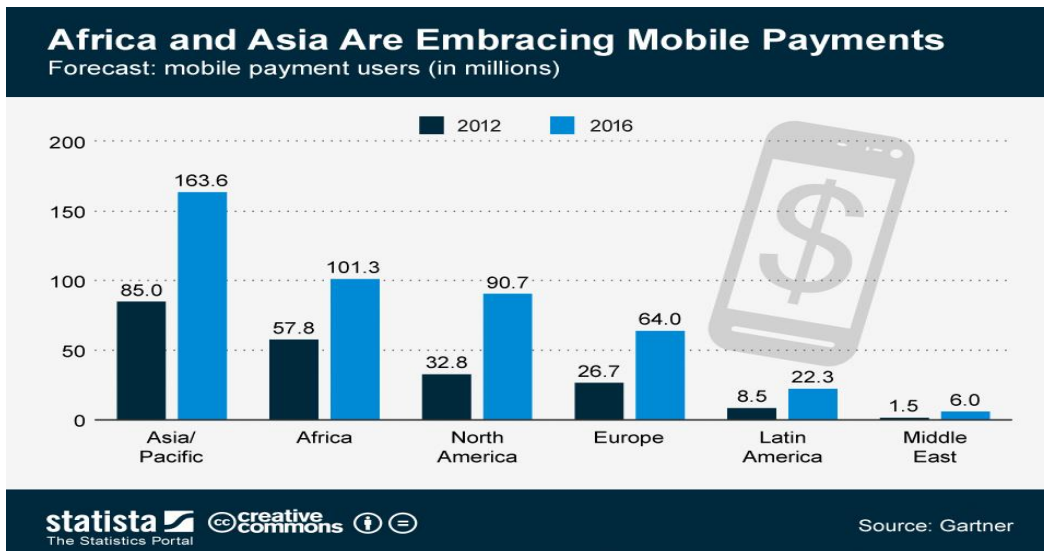


Figura 6: Paypal on Digital Wallet Market

El uso de los dispositivos digitales para transacciones comerciales es cada vez más habitual:





**Figura 7: África and Asia Mobile Payments**

Conocer al cliente mejor.

Los bancos realizan un cuestionario a los clientes en sus primeras visitas en donde se disecciona el perfil personal, familiar, patrimonial y sus objetivos financieros y, en base a ellos se aconsejan unos productos financieros.

Pero esa información, posteriormente, debe revisarse porque en nuestra vida estamos en continua evolución. Existen eventos que pueden hacer cambiar los planes y objetivos iniciales.

Los bancos necesitan de la intencionalidad del cliente en notificar dichos eventos y muchas veces no disponen de los medios para recopilar dicha información o puede ser considerada incluso un ataque a la privacidad. Pero en Facebook la actualización es inmediata

Por tanto la evolución se encuentra en la relación con el cliente, en patentar plataformas que aconsejen sobre los productos financieros de una manera más activa y adaptable a los cambios en nuestra vida.

Por ello la evolución del *Big data* financiero viene motivado por la aparición de una nueva moneda digital que sea un medio disponible, informativo y adaptable.



## 2.5.2 Big Data en las compañías aseguradoras

La función básica de las compañías aseguradoras es la generación de beneficios asumiendo riesgos de terceros. Unos riesgos que se ven drásticamente minimizados mediante el uso del **Big Data Analytics**: el acceso a datos relevantes que aporten información sobre salud, siniestralidad y morosidad entre distintos sectores de su público

Según el último estudio de *Power Data* las estrategias *Big Data* más efectivas para el sector asegurador.

### 1. Los modelos analíticos y las reclamaciones

Los modelos predictivos pueden detectar la propensión de un asegurado a contratar a un abogado a la hora de realizar una reclamación tras un siniestro. Disponer de esta información ayuda a las compañías de seguros a manejar el porcentaje de casos que podrían acabar gestionados por juristas. Es necesario apoyarse en soluciones predictivas, que clasifiquen las reclamaciones en base al impacto financiero y servicio al cliente.

### 2. Algoritmos de detección de fraude: Análisis de cobro y modelado de morosidad

La creación de un modelo de análisis y predicción de la morosidad para prevenir las pérdidas ocasionadas por reclamaciones fraudulentas. Gracias a estos modelos se pueden predecir, identificar y clasificar las cuentas más propensas a causar impagos.

Emplear un sistema de clasificación que permita identificar a solicitantes de alto riesgo de dos formas diferentes:

### 3. Análisis Cross, gestión de la fidelidad y ciclos de vida

Cross-selling: Táctica mediante la cual un vendedor intenta vender productos complementarios a los que consume el cliente. Es el algoritmo usado por Amazon para proponerte nuevas compras según tus compras anteriores o compras de otros clientes con perfil similar.

### 4. Identificación de los clientes clave.

Se pueden desarrollar modelos predictivos sobre el grado de lealtad de los clientes. Estas técnicas avanzadas ayudan a las compañías de seguros a identificar a sus

clientes más valiosos. A su vez se puede obtener datos de los clientes que tienen mayor probabilidad de fugarse a otra compañía.

### 2.5.3 Big Data en Ciencia: El científico de datos

“Antes hacíamos ciencia estudiando un gen, que al extenderlo podría contener mil palabras; ahora trabajamos con el genoma completo, que equivale a 3.000 millones de letras. Es un cambio del día a la noche”,

**“Silvia Restrepo, decana de la Facultad de Ciencias de la Universidad de los Andes”.**

El aumento del volumen de datos es producto de la “ciencia integrativa”, tendencia de cruzar distintas disciplinas en una investigación, así como de unir diferentes experimentos para explicar una hipótesis.

Según el último estudio de Universia sobre las profesiones del futuro es inminente el nacimiento de un profesional que conozca y genere un uso a esta información del Big Data: el científico de datos

Actualmente se lleva a cabo el proyecto de ‘*Big data*’ más grande de la historia: la búsqueda del origen del Universo a través del colisionado de partículas en el laboratorio CERN, donde en un solo segundo se acumula en información lo equivalente a 10.000 enciclopedias británicas, según **Bernardo Gómez, investigador de los Andes.**

El científico de datos analiza, interpreta y comunica las nuevas tendencias en el área y las traduce a la empresa para que ésta haga uso de ellas y adapte sus productos y servicios, y cree nuevas oportunidades de negocio. *Google*, por ejemplo, tiene 600 personas dedicadas al estudio del *Big Data*.

El matemático **José Antonio Guerrero** es considerado el mejor científico de datos del mundo. Para Guerrero, uno de los ejemplos más extraordinarios de la ciencia de datos es **Netflix**, que en 2006 puso a disposición de la comunidad científica de datos una base de más de **100 millones de recomendaciones de películas de 500.000 usuarios** y propuso que el primer equipo en crear un algoritmo que mejorara su sistema en un 10% recibiría 1 millón de dólares americanos. La labor tomó 3 años y 40.000 equipos, pero hoy en día representa un modelo para la mayoría de las tiendas online.

## 2.5.4 Big Data en la medicina de las 4P: personalizada, predictiva, preventiva y participativa”.

Historias clínicas e información genética de pacientes, resultados de ensayos clínicos (radiográfica), sensores que monitorizan parámetros biométricos, apps de salud... Estos datos son **desestructurados** y no pueden gestionarse con bases de datos tradicionales.

Según el último artículo de la *Fundación Telefónica “La medicina del futuro pasa por big data” de Julián García Barbosa*, el gran reto está en ser capaces de integrar todos esos datos de fuentes tan heterogéneas y analizarlos de forma conjunta:

**-Personalización:** Gracias a las tecnologías de *Big data* y los avances en la medicina genómica se podrá ofrecer a cada paciente la terapia más adecuada, podremos pasar de una medicina enfocada a poblaciones a otra basada en el individuo.

**-Predicción:** Gracias al análisis de datos procedentes de los equipos que monitorizan a se podría detectar la existencia de una posible infección mucho antes de que aparecieran los primeros síntomas.

**-Prevención:** La actriz Angelina Jolie se sometió a una doble mastectomía y tomó esa decisión por una prueba de secuenciación de ADN que reveló que tenía una mutación del gen BRCA1, que aumentaba en un 87 por ciento las probabilidades de desarrollar cáncer de mama. Gracias al diagnóstico temprano y a la intervención llevada a cabo, la actriz ha prevenido esta enfermedad y no la sufrirá en el futuro.

**-Participación:** Modelo de atención médica en el que se destaca el **papel proactivo de un nuevo tipo de “paciente digital”** que utiliza Internet para conseguir información sobre la enfermedad que padece y comparte experiencias en distintos foros con otros individuos con enfermedades o síntomas parecidos.

## 2.5.5 Big Data en el deporte: El hombre contra la máquina.

La información es clave en el deporte de alta competición como la *Fórmula 1* donde los equipos ya examinan al detalle cualquier dato relativo a las diferentes partes del coche, al piloto, al circuito y a las condiciones atmosféricas. Un ejemplo de actualidad lo encontramos en la selección alemana de fútbol, actual campeona del mundo, que

anunció el uso de herramientas de análisis de datos para definir y concretar la estrategia que le ha llevado al triunfo.

*Adidas* ha creado Micoach Elite System, un chip que se coloca en las camisetas de los jugadores, monitorizando su velocidad, su resistencia física y otros parámetros. En total, envía más de 200 registros de datos en tiempo real. El entrenador podrá ver toda esta información calibrando cuál es su estado físico e incluso psicológico.

## 2.6 Big Data Nuevo paradigma de la Educación

Según Rosa Ana Rodríguez, Doctora en Ciencias de la Educación el modelo educativo está desfasado para la era digital. El 80 por ciento de los empleos más demandados en Europa y Estados Unidos requiere de un buen dominio matemático y técnico, y en España sólo trece de cada mil jóvenes terminan itinerarios de este tipo.

### Tendencias en Educación: la integración de las TIC de forma holística

En Estados Unidos las tendencias se dirigen al fomento de las competencias STEM (Ciencia, Tecnología, Ingeniería y Matemáticas). En España las comunidades autónomas establecen diferentes líneas de actuación que tienen como punto de partida la integración de las TIC:

**El modelo pedagógico:** una de las tendencias metodológicas que aparecen en el *Informe Horizon 2014: Enseñanza Superior es la integración del aprendizaje en línea y colaborativo*, la posibilidad de que los estudiantes sean también creadores de contenido. Surge **ScienceLab**, es un modelo alemán de aprendizaje, que pretende aprovechar la necesidad innata de entender el mundo que tienen los niños de entre 4 y 10 años. Consiste en que el niño sea dueño de su propio proceso de aprendizaje, a partir de las preguntas que le van surgiendo. Adquieren los conceptos clave, siguiendo el método científico de pregunta, hipótesis, experimentación, reformulación y conclusiones.

### 2.6.2 Los cursos MOOC: España a la cabeza de Europa en creación de ellos.

Según **Kenneth Cukier** en *“La revolución de los datos masivos”* describe cómo el universo educativo puede beneficiarse de esos datos. **Kenneth Cukier** es un periodista estadounidense, gurú en *Big Data* y en posesión de múltiples publicaciones y conferencias.

*“Lo que ha sucedido en Estados Unidos, con los cursos en línea, llamados MOOCs, cursos masivos abiertos en línea, es que los profesores pueden ver cuando los estudiantes están viendo sus cursos, y cuando se detienen, cuando releen una lección. Y un profesor de Stanford se dio cuenta de que hacia la lección siete u ocho todos los estudiantes regresaron a la lección número tres. Esa lección era una clase de repaso de*

matemáticas y mostraba que a medida que los estudiantes avanzaban más en el curso, estaban menos seguros de sus bases en matemáticas. Normalmente, un profesor no tiene porqué saber que la clase se está quedando atrás, pero de repente, el profesor podía ver esto en los datos y podía aprender dos cosas. En primer lugar, que debía preparar mejor a sus alumnos. Y en segundo, que debía insistir más en esa dificultad en particular ya que los estudiantes se detenían y regresaban a esa lección.”

Kenneth Cukier

Kenneth Cukier anuncia una educación que se adapta a las necesidades de cada alumno:

“Tenemos que proponer a nuestros hijos otro sistema educativo ya que éste fue concebido en una época diferente, en la era industrial, mecanicista, basada en una línea de montaje. Ahora se puede adaptar a las recomendaciones de *Amazon* y *Google* que se ajustan exactamente a nuestros intereses. Necesitamos una educación que se adapte a nuestras necesidades y esa es la mejor manera de aprender.”

El informe de la **Comisión Europea “European MOOC Scoreboard”** sitúa a España como líder en el ranking de producción de MOOCs con más de 151 cursos de un total de 432 en toda Europa.



Figura 8: Distribucion de los MOOCs fuente “Comision European MOOC Scoreboard”

### 2.6.3 Estrategias de innovación Educativa.

En la *Universidad Estatal de Arizona* un software recoge toda la información y datos de cada estudiante. Notas, las habilidades y dificultades, e incluso sus vacilaciones al usar el mouse. Entonces, unos algoritmos predictivos, comparan las estadísticas con los datos recogidos entre decenas de miles de estudiantes.

“Es un método muy personalizado que se adapta a las necesidades de cada alumno, explica *Irene Bloom profesora de la Universidad Estatal de Arizona*. Por ejemplo un estudiante puede avanzar sin problemas en una asignatura, mientras que otro necesitará más trabajo y el sistema se adapta en función de cada cual.”

*Art Blakemore, vicerrector de la Universidad Estatal de Arizona:*

“Nuestro objetivo principal es conseguir que los estudiantes logren terminar sus estudios. Además es un buen sistema para retener a los estudiantes que seguirán pagando sus matrículas. La tasa de éxito de la Universidad Estatal de Arizona ha aumentado un 13% y el abandono se redujo en un 54% durante el primer año. “

- Wepack

En España **Telefónica Educación Digital** ha impulsado *Wepack*. *Wepack* ofrece una solución global para todas las etapas escolares y tiene en cuenta varios de los pilares de la innovación digital:

Entorno Virtual de Aprendizaje, la plataforma *WEclass* para gestionar los procesos de enseñanza y aprendizaje. **Scolartic**, red social para familias y docentes especializados con el fin de intercambiar información.

Soluciones integradas en el aula a través de la conexión de dispositivos móviles de forma segura y entornos cloud para acceso desde Smartphone o table.

Existen diferentes iniciativas para entrenar las habilidades cognitivas en el ámbito sanitario para por ejemplo ayudar a las personas con déficit de atención o alteraciones de la memoria. *Wepack Infantium*, utiliza la tecnología inspirada en el cerebro con aplicación directa en la personalización del aprendizaje. Permite crear itinerarios individuales con la integración de contenidos según las capacidades cognitivas de los estudiantes de Educación Infantil, su estilo de vida y su velocidad de

aprendizaje. Gracias al *Big data* se registran las secuencias de aprendizaje del alumno y cómo aprende su cerebro uniendo la tecnología de los datos y la neuro educación.

- **Punto Neutro educativo: Plataforma única de recursos educativos digitales.**

Plataforma de recursos digitales para el aprendizaje cuyo objetivo es hacer accesible la cultura independientemente del poder adquisitivo. Se enmarca en **“Plan de cultura digital en la escuela”** del *Ministerio de Educación* y ha sido adjudicado a *Telefónica*.

Punto Neutro es un sistema de información orientado al tratamiento y administración de recursos digitales educativos en un catálogo bajo estándares reconocidos y una plataforma segura de compra - venta.

### Beneficios del Punto Neutro:

Centros educativos y docentes: Permite consultar toda la oferta de los proveedores de contenidos educativos.

Alumnos y familias: Disponen de una única plataforma para la compra del material digital.

Editores y desarrolladores de contenidos: Poseen un único escaparate para.



Figura 9: Explicación Punto Neutro fuente Ministerio de Educación.





## 3. DISEÑO DE UNA ARQUITECTURA BIG DATA

---

Existen múltiples implementaciones de la arquitectura *Big Data* tradicionales basadas en Business Intelligence:

### 3.1 Fase 1: Recolección de datos.

En capítulos anteriores hemos definido la enorme variedad de fuentes de información en los sistemas *Big Data*. La recolección de datos se clasifica según sean:

- **Streaming:** Obtención en tiempo real de los datos. Utilizado en sistemas de seguridad como por ejemplo aviación comercial y sistemas de seguridad que precisan de respuestas inmediatas.
- **Batch:** Obtención de datos a intervalos de tiempo. Se suele utilizar en bases de datos cuando la información por ejemplo solo se analiza finalizada la jornada o cuando se quieren analizar volúmenes de datos en periodos determinados de actividad.

### 3.2 FASE 2 ALMACENAMIENTO: HADOOP DISTRIBUTED FILE SYSTEM (HDFS)

HDFS es el sistema que usa Hadoop para el almacenamiento de archivos. HDFS permite conectar nodos para formar clústeres, en los nodos se almacena la información particionada. Cada bloque se replica en máquinas separadas físicamente aumentando la tolerancia a fallos. HDFS permite aumentar la eficiencia gracias a ejecutar los cálculos en la máquina donde residen los datos en lugar de donde se realizan las peticiones.

HDFS tiene dos tipos de nodos, el maestro *NameNodes* y el esclavo *DataNodes*:

***Namenodes***: controla las peticiones a los archivos y es el encargado de distribuir la replicación de los bloques entre los diferentes datanodes.

***Datanodes***: realizan operaciones de abrir, cerrar y renombrar archivos, se encargan de almacenar y recuperar los datos. Los *datanodes* crean, eliminan y replican bloques de datos según las órdenes del *Namenode*. Los *datanodes* periódicamente informan al *Namenode* las listas de bloques se están almacenando. Por cada grupo de Datanode existe un Namenode asociado.

El protocolo de comunicación de HDFS es TCP/IP. HDFS se conecta a un puerto TCP abierto en el namenode, y luego se comunica con el namenode usando el protocolo RPC (Remote Protocol Call).

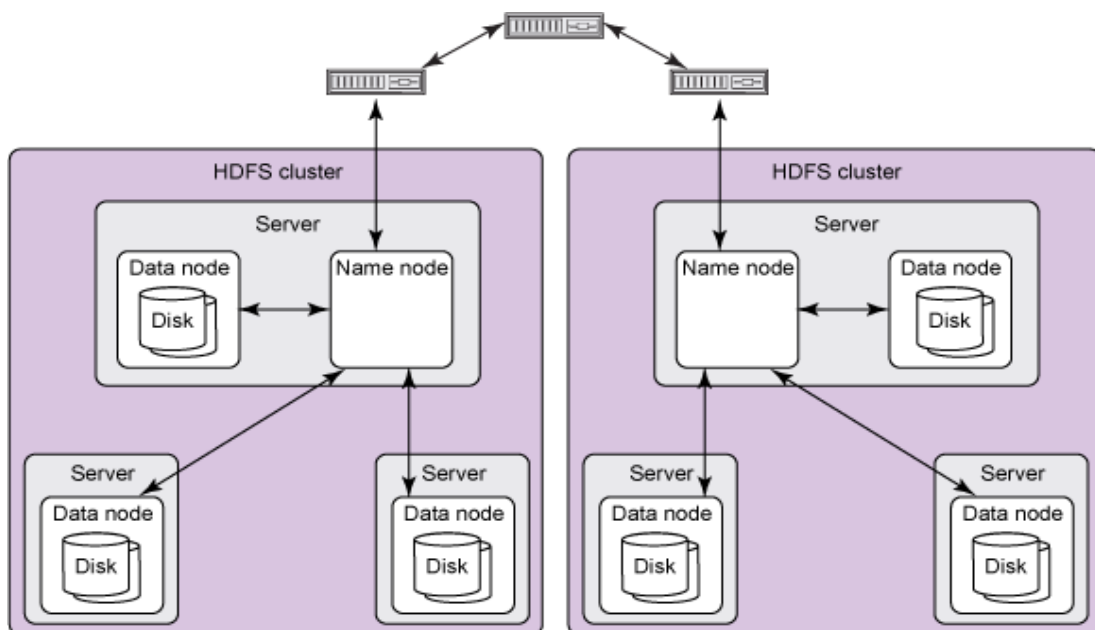


Figura 10: Arquitectura HDFS fuente "An introduction to the Hadoop Distributed File System" IBM

### 3.2.1 BASES DE DATOS RELACIONALES; SQL-NoSQL

**SQL** (*Structured Query Language*) es un grupo de lenguajes para trabajar con bases de datos. Los lenguajes SQL hacen uso de consultas sencillas pero el principal problema de los sistemas SQL es que necesitan transcribir cada sentencia para poder ser ejecutada por tanto muchas peticiones pueden ralentizar el sistema.

```
SELECT p.Nombre, p.PrimerApellido, c.Vehiculo
FROM Personas p, Contratos c
WHERE p.Teléfono == c.Teléfono
```

Tabla 3: Ejemplo consulta SQL

**NoSQL** aparece con la web 2.0, hasta ese momento sólo subían contenido a Internet las empresas pero con la llegada de las redes sociales como Facebook, Twitter o Youtube, cualquier usuario podía subir contenido, provocando un crecimiento exponencial del volumen de datos.

Las bases de datos NoSQL son sistemas de almacenamiento de información que no utilizan una estructura de datos en forma de tabla donde se van almacenando los datos. *Se ejecutan en máquinas de coste y recursos reducidos y la información puede estar compartida en varias máquinas.*

#### NoSQL: Modelo Clave-Valor

La información (llamada valor) es almacenada e identificada por una única clave.

En un sistema clave-valor tenemos contenedores que pueden almacenar tantas parejas de clave-valor como queramos. Cada contenedor puede tener datos de la misma naturaleza (por ejemplo productos, pedidos, clientes, etc.) o totalmente diferente (un contenedor por cliente).

El único modo de realizar consultas es a partir de la clave para extraer información, el sistema solo reconoce claves y con ellas extrae valores nunca al revés.

Este modelo es utilizado para almacenar información sobre la sesión de un usuario y comercio electrónico donde se puede asociar una clave numérica o usuario y almacenar en un mismo contenedor todos los datos relacionados con el usuario. Este tipo de sistemas solo sirve para el almacenamiento y extracción de información.

Algunos de las implementaciones más conocidas son *SimpleDB de Amazon*, o *BigTable de Google*.

### Horizontal

ID	Usuario	Nombre	Email	Fecha
1	Isma	Ismael Herrera	<a href="mailto:Ismael@gmail.com">Ismael@gmail.com</a>	25/01/1988
2	Juan	Juan Simon	<a href="mailto:Juan.simon@hotmail.com">Juan.simon@hotmail.com</a>	null
3	María	Maria Gomez	<a href="mailto:maria@gmail.com">maria@gmail.com</a>	25/03/1990

Tabla 4: Ejemplo sistema de identificación de usuarios. Horizontal.

En el primer contenedor la clave es el nombre de usuario y el valor la contraseña.

### Vertical

ID USUARIO	CLAVE	VALOR
1	Nombre	Ismael Herrera
1	Email	<a href="mailto:Ismael@gmail.com">Ismael@gmail.com</a>
1	Fecha	25011988
2	Nombre	Juan Simon
2	Email	<a href="mailto:Juan.simon@hotmail.com">Juan.simon@hotmail.com</a>
3	Nombre	Maria Gomez
3	Email	<a href="mailto:Maria@gmail.com">Maria@gmail.com</a>
3	Fecha	25031990

Tabla 5: Ejemplo sistema de identificación de usuarios. Vertical.

## Modelo de datos documental

La información es almacenada en formato ASCII en estándares como JSON (Javascript Object Notation). La información almacenada en un documento está formado por pares de clave y valor. Las bases de datos documentales están libres de esquemas y de estructuras ya que son datos almacenados en un documento, la información no se divide en columnas y filas, si no en documentos con libertad de información. Cada objeto dentro del documento contiene información sobre otros sub-objetos que guardan asociación con el principal dentro de un mismo documento.

La inclusión, modificación y extracción o minería de datos se realiza mediante framework o API que permiten buscar objetos por palabra clave, rango o cualquier otra condición regularmente utilizada en bases de datos relacionales.

Dos de las bases de datos más extendidas son MongoDB y CouchDB.

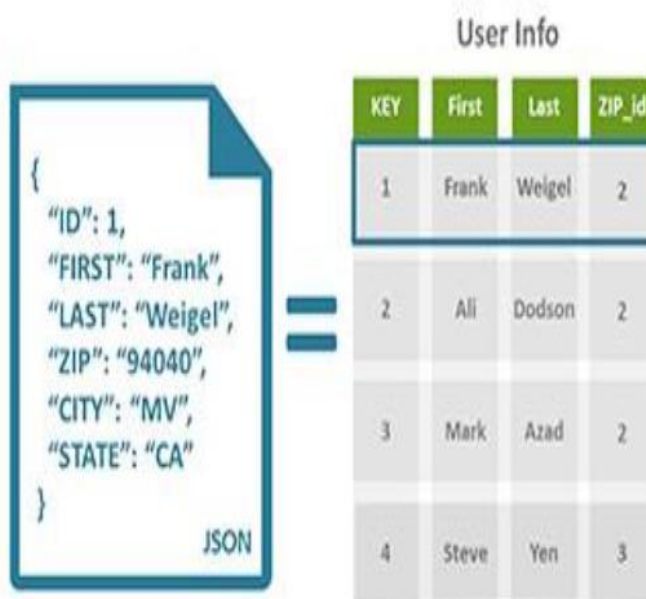


Figura 11: Ejemplo de Base de datos documentales fuente "IBM Developers Work"

## El Modelo de Grafos

La información se representa como nodos de un grafo. En redes sociales son utilizados para imitar las relaciones, los gustos y las preferencias de los miembros de una comunidad o los patrones de comportamientos entre miembros con gustos afines.

Cada individuo es representado por un nodo y cada nodo está relacionado con los demás mediante aristas que almacenan información sobre quien es amigo de quien y quien es amigo de los amigos de mis amigos.

En adición tenemos otros nodos que guardan información sobre las tiendas visitadas y otros sobre los artículos adquiridos. Estos nuevos nodos se relacionan con los individuos de forma que podemos saber que tiendas visita la gente y que suelen comprar ellos y sus amigos. Gracias al grafo generado por los nodos y sus relaciones es posible también realizar estudios sobre si los individuos que tienen alguna relación entre si suelen visitar las mismas tiendas, si los patrones de compras de cada uno de los individuos y sus amigos se parecen y como compran según su nivel adquisitivo entre otros.

Las bases de datos más conocidas y utilizadas en los mercados son Neo4j y HyperGraphDB.

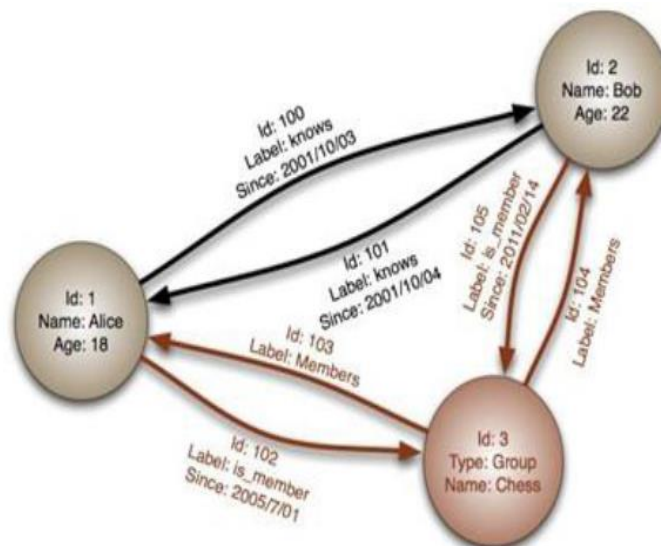


Figura 12: Ejemplo modelo Grafos fuente “Fundación Telefónica”.

### 3.3 Análisis de datos: MAPREDUCE

El concepto MapReduce aparece por primer vez en 2004 en el paper de Google “MapReduce: Simplified Data Processing on Large Clusters”. El objetivo principal de MapReduce era permitir la computación paralela sobre grandes cantidades de datos.

#### MapReduce consta de 2 funciones:

- Map toma como entrada un par (clave,valor) y devuelve una lista de pares (clave2,valor2)

La función es aplicada en paralelo para cada ítem produciéndose una lista de pares (k2,v2) por cada llamada.

$Map(k1,v1) \rightarrow list(k2,v2)$

```
function wordMap() {  
  
    // Busca palabras diferentes  
  
    var palabras = this.text.match(/\w+g);  
  
    if (palabras == null) {  
  
        return;  
  
    }  
  
    for (var i = 0; i < palabras.length; i++) {  
  
        // Salida cada palabra con contador "1" asociado.  
  
        emit(words[i], {count: 1});  
  
    }  
  
}
```

Después el framework de MapReduce junta todos los pares con la misma clave de todas las listas y los agrupa, creando un grupo por cada una de las diferentes claves generadas. El nodo master toma el input, lo divide en problemas de menor identidad, y los distribuye a los denominados worker nodes. El worker node procesa el problema y pasa la respuesta al nodo maestro.



**Reduce:** La función reduce es aplicada en paralelo para cada grupo, produciendo una colección de valores para cada dominio:

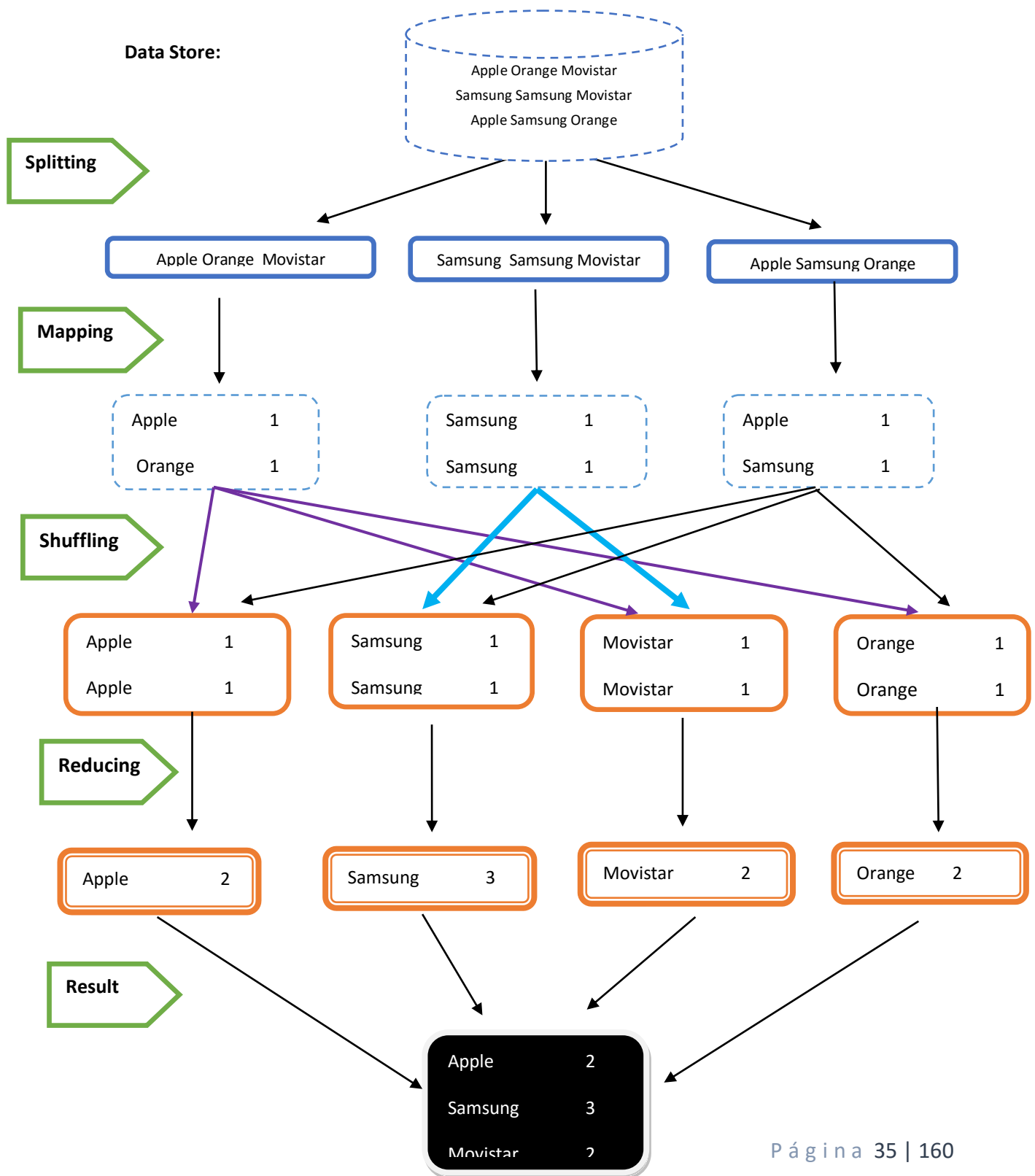
*Reduce(k2,list(v2))→list(v3)*

El framework MapReduce transforma una lista de pares (clave, valor) en una lista de valores.

```
function wordReduce(key, values)
    var total = 0;
    for (var i = 0; i < values.length; i++) {
        total += values[i].count;
    }
    return {count: total};
}
```

Ejemplo: Base de datos de marcas de Smartphone de usuarios.

Figura 13: Fases del algoritmo MapReduce fuente "IBM Developers Work Hadoop"



### 3.3 interfaz de usuario: VISUALIZACIÓN

Las herramientas para facilitar la visualización gráfica de los datos como las infografías están ganando cada vez más popularidad como recurso educativo debido a su capacidad para simplificar y hacer más atractiva la información. A base de mapas, iconos, gráficos, datos y pequeños textos nos permiten entender los datos de una manera visual e interpretando la relación entre ellos. Existen múltiples herramientas para la visualización de datos.

#### 3.3.1 Infografías con Easel.ly:

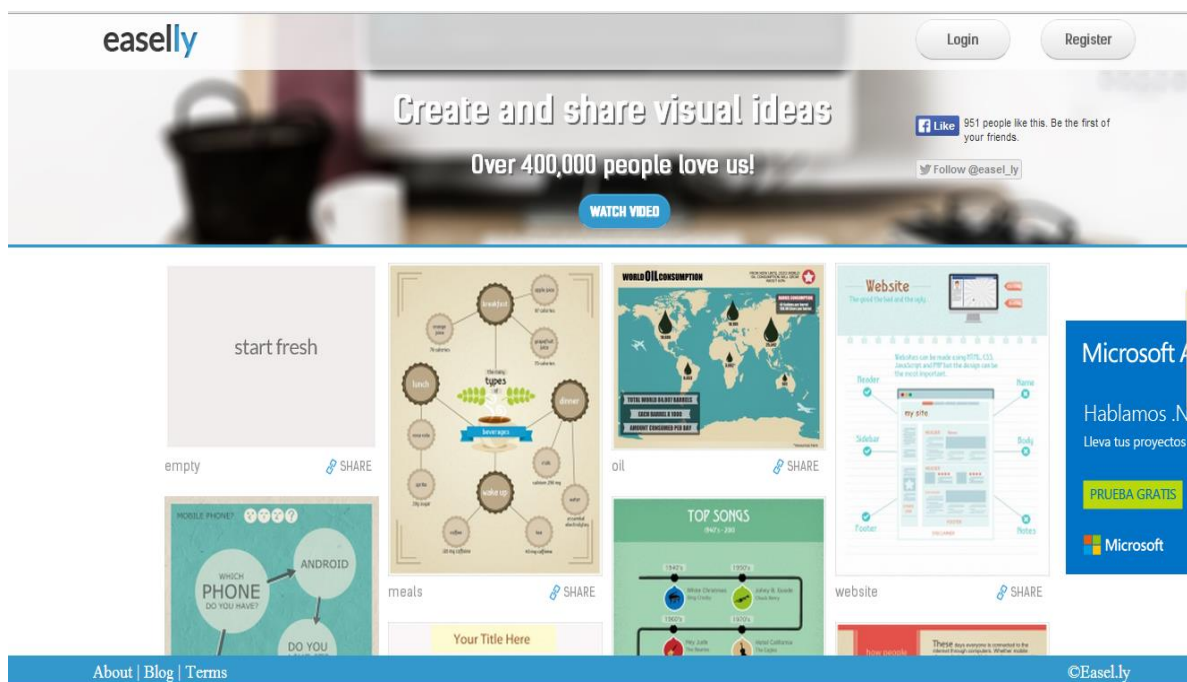


Figura 14: Interfaz de usuario Easel.ly.

Aparecerá una pequeña **barra de herramientas del editor**, que nos permitirá realizar las siguientes acciones propias de edición:



Figura 15: Barra de herramientas de Easel.ly.

- **Seleccionar el tema: Vhemes**

A partir de una plantilla o desde cero, se puede crear una infografía con esta aplicación que es del tipo Wysiwyg.

- **Añadir objetos, formas o textos**

Objects: Todos los objetos que pueden agregar están clasificados en categorías y además también nos permite usar nuestras imágenes.

Al desplazar los objetos por el lienzo, aparece líneas guía de posición del objeto, que son muy útiles para conseguir que todos los elementos queden bien alineados.

- **Backgrounds:** opción de cambiar la forma del fondo o su color.
- **Shapes:** incluye flechas, bocadillos, formas redondeadas o cuadradas, etc.
- **Text:** da acceso a un pequeño editor de texto mediante el que puedes modificar la tipografía, el tamaño o el color de los textos.

Las infografías pueden ser exportadas como archivo PNG, JPG, PDG, y SVG. Los usuarios pueden registrarse en Easel.ly para poder guardar sus propias infografías.

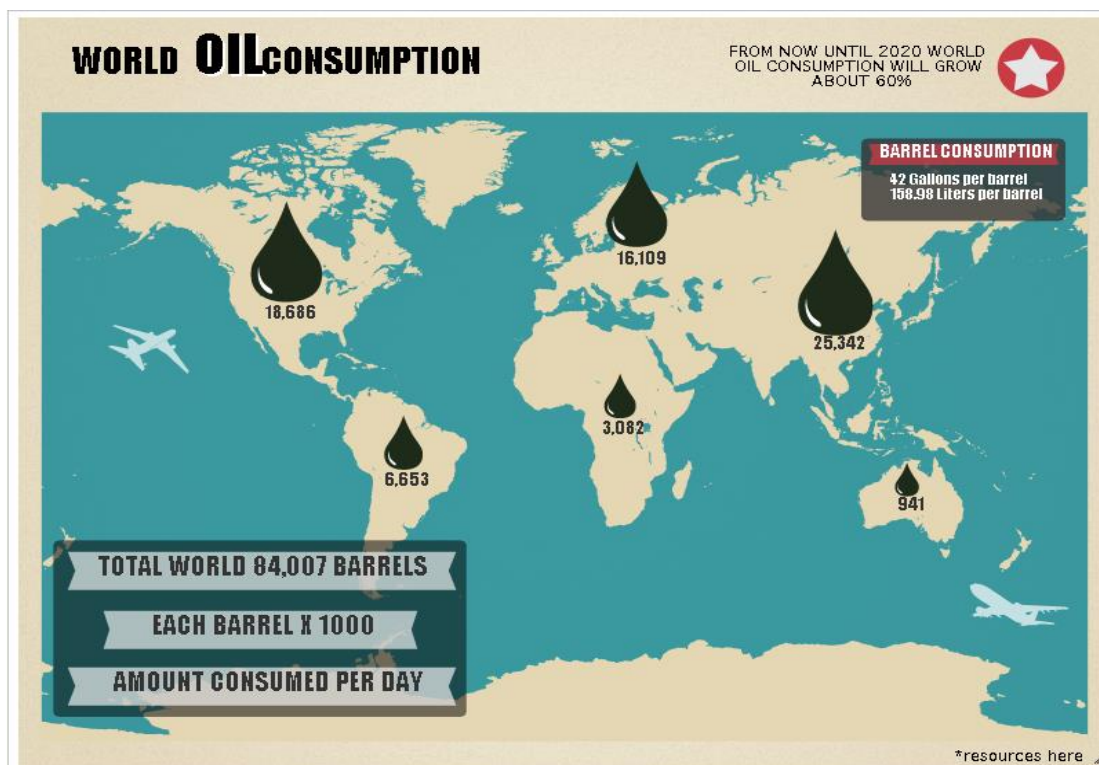


Figura 16. Ejemplo de infografía obtenido con Easel.ly sobre consumo de carburante.

### 3.3.2 Infografías con Piktochart

Piktochart permite el uso de un amplio número de plantillas para infografías, informes, banners y presentaciones. El sistema funciona a través de bloques, lo que disminuye la flexibilidad de la plantilla.



Figura 17: Piktochart interfaz de usuario.

Cuenta con un menú similar a WordPress para hacer cambios de forma sencilla.



Figura 18: Piktochart herramientas de edición.

Tiene la opción de insertar vídeos, mapas con datos demográficos y gráficos. Piktochart te permite exportar tu infografía en formato PNG, JPG y también en PDF para insertar en nuestra página web o podemos publicarlo directamente en nuestros perfiles en redes sociales.

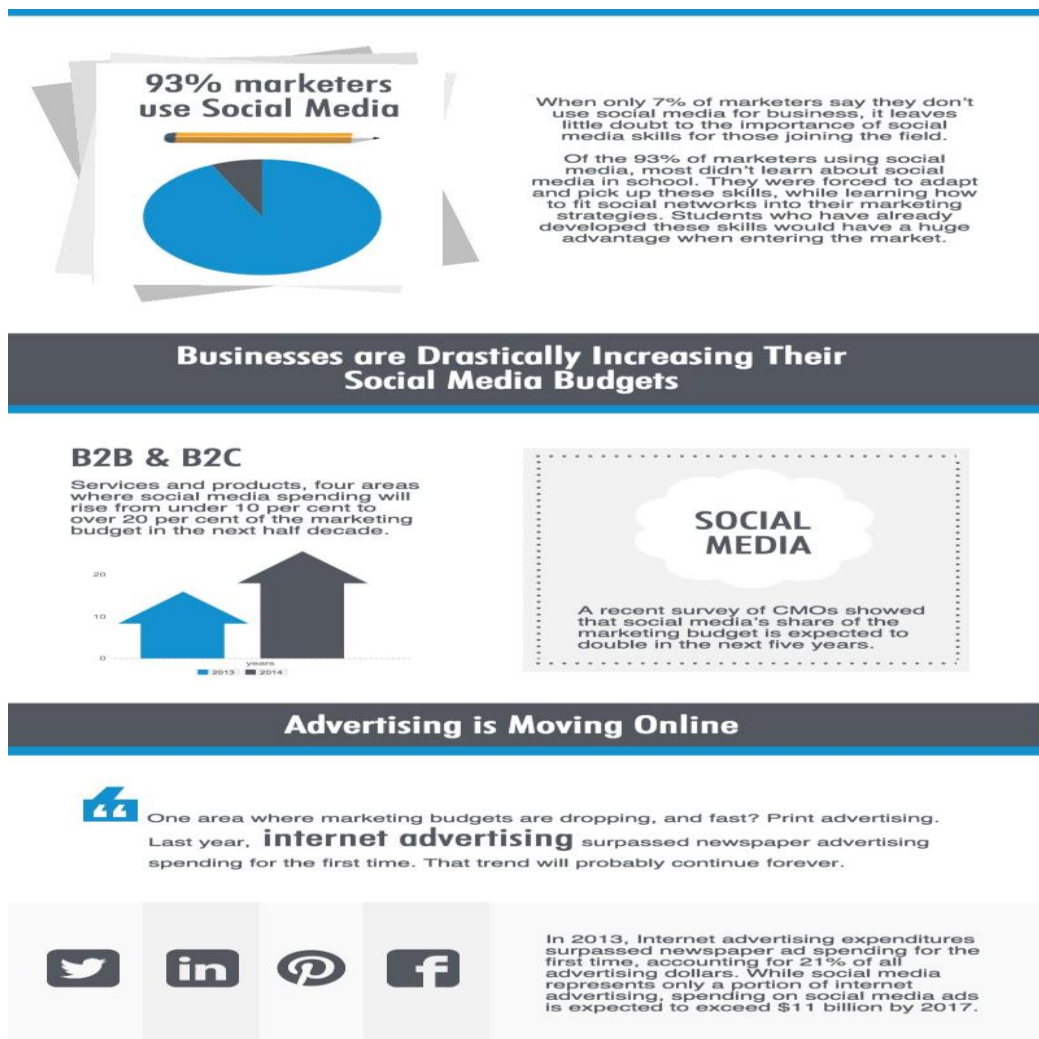


Figura 19: Piktochart ejemplo de infografía sobre Social Media.

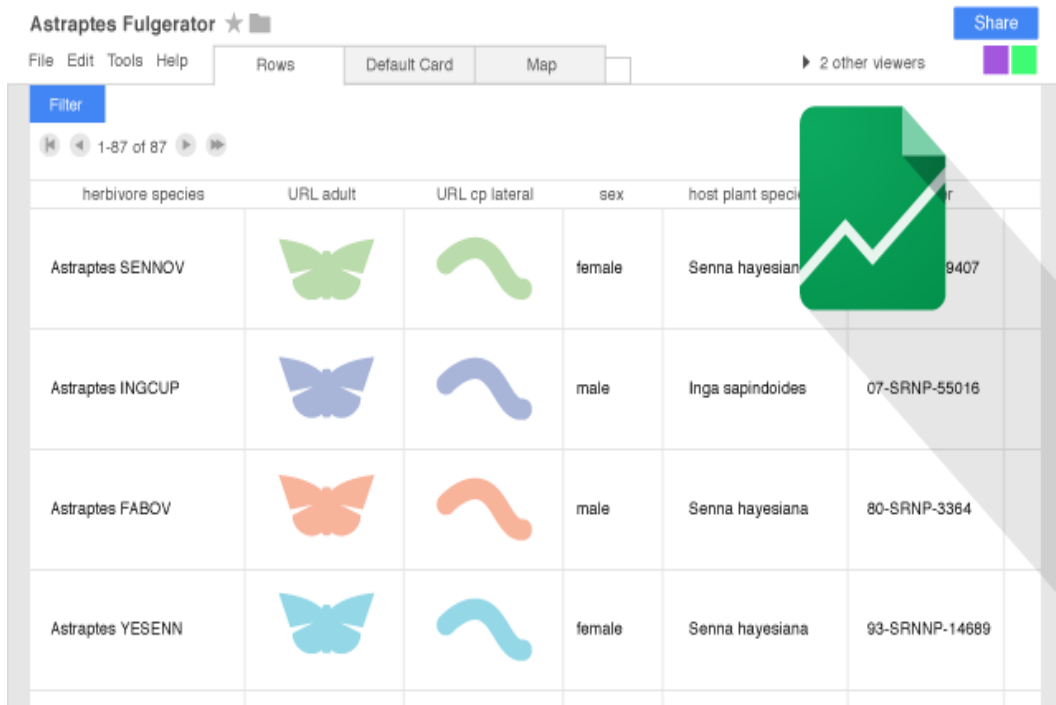
### 3.3.3 Google Fusion Tables:

*Google Fusion Tables* es un servicio web de gestión de datos que permite generar mapas y detallados gráficos. Puede ser usado de forma colaborativa para trabajar con datos públicos o subidos por nosotros.

Los datos se almacenan en varias tablas que los usuarios de Internet pueden ver y descargar. El sitio web se inauguró en 2009 anunciado por *Alon Halevy* y *Shapley Rebecca*.

El servicio web proporciona un medio para visualizar los datos con gráficos circulares, gráficos de barras, diagramas de dispersión y líneas de tiempo; así como mapas geográficos basados en *Google Maps*. Los datos se exportan en un formato de archivo y los valores son separados por comas.

El papel de **Google Fusion Tables** fue descrito como importante para la investigación y la representación de datos en 2010.2



The screenshot shows the Google Fusion Tables interface for a table named "Astraptles Fulgerator". The table has columns for "herbivore species", "URL adult", "URL cp lateral", "sex", "host plant species", and "id". The data rows are as follows:









herbivore species	URL adult	URL cp lateral	sex	host plant species	id
Astraptles SENNOV			female	Senna hayesian	9407
Astraptles INGCUP			male	Inga sapindoides	07-SRNP-55016
Astraptles FABOV			male	Senna hayesiana	80-SRNP-3364
Astraptles YESENN			female	Senna hayesiana	93-SRNP-14689

Figura 20: Ejemplo de Fusion Tables de *Google* con tipo de mariposa fuente *Google*.



## 4 SISTEMA HADOOP



*Hadoop* surgió como iniciativa de software libre a raíz de la publicación del *paper* [Bigtable: A Distributed Storage System for Structured Data](#) de *Google* sobre sus sistemas de archivo, su herramienta de mapas y el sistema Bigtable.

BigTable es un mapa distribuido en: filas, columnas y tiempo.

Las claves de las filas de BigTable son cadenas de tamaño variable y no se pueden producir dos operaciones sobre una misma fila. Las filas son particionadas en “tablets” de unos 200 Mbytes, que se dispersan entre varios nodos. Cada nodo almacena 100 tabletas, mediante el sistema 'Google File System'.

Las claves de las columnas forman las “familias de columnas”, a las que se les puede asignar un valor determinado para cada fila.

Estas celdas disponen de versiones temporales de sus valores. Todo se muestra en la siguiente imagen, obtenida del paper:

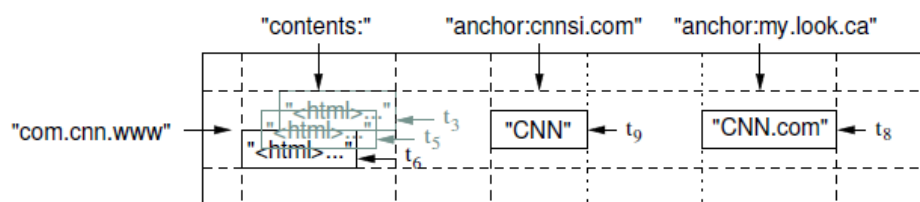


Figure 1: A slice of an example table that stores Web pages. The row name is a reversed URL. The *contents* column family contains the page contents, and the *anchor* column family contains the text of any anchors that reference the page. CNN's home page is referenced by both the Sports Illustrated and the MY-look home pages, so the row contains columns named *anchor:cnnsi.com* and *anchor:my.look.ca*. Each anchor cell has one version; the contents column has three versions, at timestamps  $t_3$ ,  $t_5$ , and  $t_6$ .



Como resultado nació *HDFS Apache, Apache MapReduce y Apache HBase*. Hadoop es capaz de almacenar toda clase de datos: estructurados, no estructurados, semiestructurados. Destaca por tener la capacidad de asegurar una alta disponibilidad y recuperación de los datos que procesa.

Alguna de las empresas y redes sociales que actualmente usan Hadoop:

- **Facebook:** Utilizan Hadoop para almacenar copias de *logs* y fuentes de datos de grandes dimensiones, así como de fuente de informes/análisis.
- **Google:** Junto con *IBM*, lleva a cabo la iniciativa **“University Initiative to Address Internet-Scale Computing Challenges”** que intenta mejorar el conocimiento de los estudiantes sobre computación paralela para adaptarse a modelos de programación como MapReduce.

**Twitter:** Utilizan Hadoop para almacenar y procesar *tweets* y ficheros de *log*.

#### 4.1 HADOOP DISTRIBUTED FILE SYSTEM (HDFS) 1.0



Idóneo para aplicaciones que realicen escritura del fichero una sola vez. De esta manera se puede conservar la coherencia de los datos y habilita su acceso rápido. Permite añadir nuevos nodos sin tener que pausar o parar los procesos.

El tamaño típico de bloque es de 64MB, por lo tanto cada archivo HDFS consiste de uno o más bloques de 64MB.

## Arquitectura

HDFS se basa en una arquitectura Maestro/Esclavo con dos tipos de nodos:

- **Nodo Máster: NameNode**

Responsable de mantener el estatus de los nodos *slave*, establece uno de ellos como nodo pasivo, que se convertirá en nodo máster en caso de urgencia.

HDFS mantiene la integridad de los datos a través de unas sumas computadas en el contenido de los archivos para la validación (*namespace*). El *namespace* se almacena en cada *namenode* en un archivo llamado *FsImage*, cuando se inicializa un *namenode*, lee el archivo *FsImage* y aplica la información de estado que se encuentra en esos archivos.

- **Nodo slave: DataNode.**

Nodos encargados de almacenar la información que se está procesando por el nodo máster. Envía de forma periódica mensajes a su *Namenode*. Si el *Namenode* no recibe ningún mensaje de algún *datanode*, marca al *datanode* como inactivo y deja de enviarle peticiones. Los datos almacenados en el *datanode* muerto ya no están disponibles por tanto se eliminan del sistema.

### **Rack**

Conjunto de nodos de datos. Cada rack tiene un switch que le permite comunicarse con los distintos racks del ecosistema, sus nodos y procesos cliente.

### **Proceso cliente**

Proceso que se lanza a petición de un *nodo máster*, ya sea para almacenamiento de archivo nuevo o recuperación de un archivo en el clúster Hadoop.

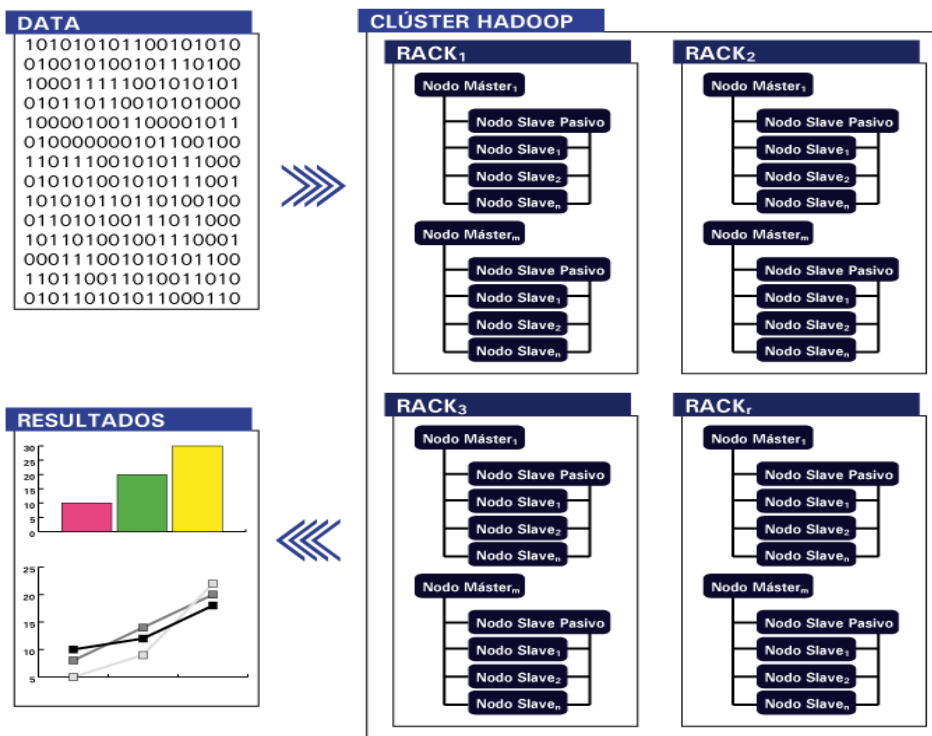


Figura 21: Ejemplo Hadoop HDFS fuente IBM Analytics What is Hadoop?

## Rebalanceo de bloques de datos

HDFS soporta rebalanceo de carga pudiendo, mover bloques de datos de un *datanode* a otro si queda poco espacio libre en el datanode o creando réplicas adicionales si se produce un aumento rápido de demanda de un determinado archivo.

### 4.1.2. MAPREDUCE 1.0

MapReduce es un modelo de programación destinado a procesar grandes cantidades de datos en paralelo y de manera distribuida sobre un cluster.

Ese procesamiento comienza por una fase llamada **Map** que se ocupa de filtrar, seleccionar, contar u ordenar un conjunto de registros.

El resultado de la fase Map es procesado en una fase posterior llamada *Reduce*. Esta fase también corre en paralelo y en general sumaria los resultados de la fase *Map*.

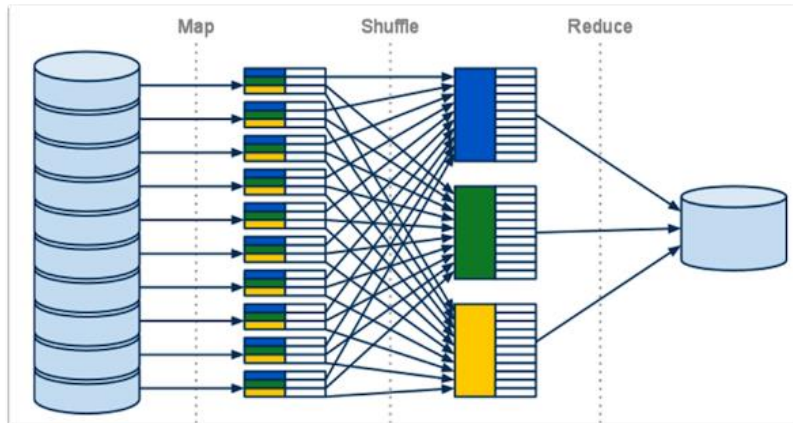


Figura 22: Arquitectura Map and Reduce fuente IBM Analytics What is Hadoop?

## ARQUITECTURA MAPREDUCE

Arquitectura maestro-esclavo con dos tipos de servicios:

- **Servicio JobTracker:** gestiona y distribuye los trabajos de MapReduce. Finalizados los procesos map envía los resultados a un nodo de trabajo para realizar la fase reduce. Permite la consulta de datos.
- **Servicio TaskTracker:** se ejecuta en un nodo DataNode de HDFS. Recibe las órdenes del JobTracker y realizar el proceso sobre el bloque de datos que contenga. Una vez finalizado el proceso se lo envía al JobTracker. Es el encargado de realizar la fase reduce.

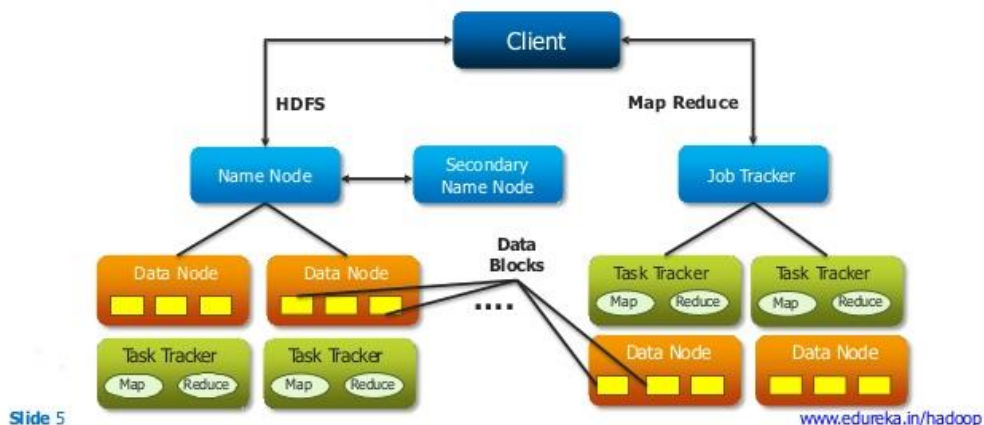


Figura 23: Workflow de la ejecución de un trabajo MapReduce en una arquitectura Hadoop.

## ETAPAS DE UN PROCESO MAPREDUCE

- **Mapeo:** los datos de entrada se encuentran en la forma de archivo o directorio y se almacenan en el sistema de archivos HDFS, en esta etapa los nodos escogidos por el JobTracker ( Mappers) procesan los datos y crean varios pequeños fragmentos de datos.

	Entrada	Salida
Mapa	$\langle k1, v1 \rangle$	lista( $\langle k2, v2 \rangle$ )

- **Shuffle/ Sort :** es el proceso donde se ordenan por claves. Los datos de salida de los map y se transfieren hacia los reducers.

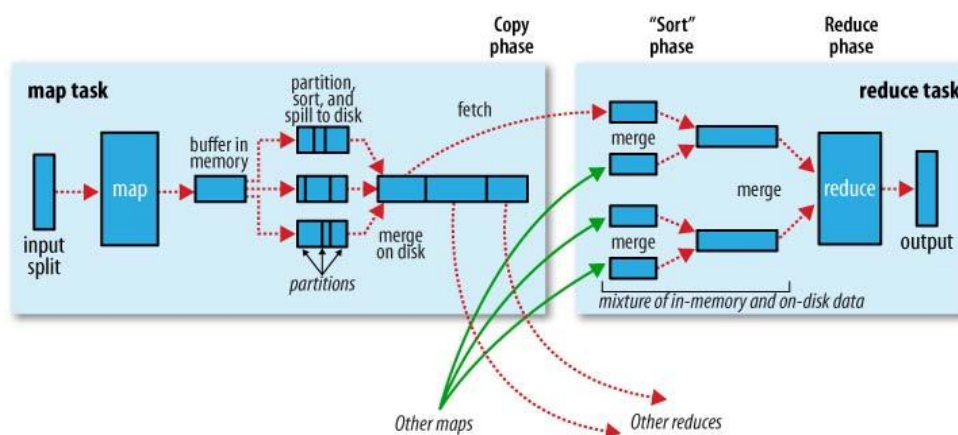


Figura 24: Workflow de la ejecución de un trabajo MapReduce en una arquitectura Hadoop.

- **Reducing:** reducción final del trabajo. La entrada del Reducer -el nodo que realiza esta función- recibe las claves y las listas de valores para estas claves. Después de un proceso de elaboración, se genera un nuevo conjunto de la producción, que se almacena en el HDFS.

	Entrada	Salida
Reduce	<k2, lista(v2)>	lista (<k3, v3>)

### MAPREDUCE: EJEMPLO TEÓRICO DE USUARIOS Y PUNTUACIONES.

Tenemos un fichero con fecha, nombre y puntuación. Queremos sacar la suma total de puntuaciones por usuario.

```
01-11-2015 Paula 11
01-11-2015 Pablo 9
01-11-2015 Juan 3
23-11-2015 Pablo 22
23-11-2015 Paula 15
15-12-2015 Pablo 32
15-12-2015 Paula 47
15-12-2015 Juan 13
01-01-2015 Pablo 2
01-01-2015 Paula 3
01-01-2015 Juan 32
```

Tabla 6: Ejemplo MapReduce, fichero inicial con fecha, nombre y puntuación de usuarios.

Map recibe las siguientes líneas, las key son el offset de la línea (en bytes), y los value son la línea de texto entera.

```
(0, 01-11-2015 Paula 11)
(22, 01-11-2015 Pablo 9)
(45, 01-11-2015 Juan 3)
(65, 23-11-2015 Pablo 22)
(86, 23-11-2015 Paula 15)
(110, 15-12-2015 Pablo 32)
(134, 15-12-2015 Paula 47)
(158, 15-12-2015 Juan 13)
(182, 01-01-2015 Pablo 2)
(205, 01-01-2015 Paula 3)
(225, 01-01-2015 Juan 32)
```

Tabla 7: Ejemplo MapReduce: Archivo Map modelo clave-valor.

- **Map**

La función *map* recoge la información de interés y omite el resto, para finalmente emitir para cada línea un par *key/value* con el nombre del jugador como key y la puntuación como value:

```
(Paula, 11)
(Pablo, 9)
(Juan, 3)
(Pablo, 22)
(Paula, 15)
(Pablo, 32)
(Paula, 47)
(Juan, 13)
(Pablo, 2)
(Paula, 3)
(Juan, 32)
```

Tabla 8: Ejemplo Mapreduce, resultado Map key/value

Esta salida del *map* va a ser tratada por el **Shuffle** and Short antes de enviar los datos a la función *reduce*. En esta fase las key se van a ordenar y los value se van a agrupar por cada key, y el resultado será la entrada de la función *reduce*, quedando de la siguiente forma:

```
(Juan, [3, 13, 32])
(Paula, [11, 15, 47, 3])
(Pablo, [9, 22, 32, 2])
```

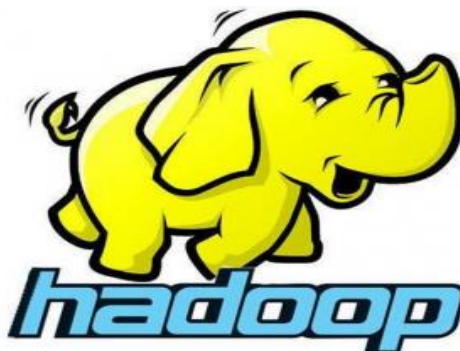
Tabla 9: Ejemplo MapReduce, proceso shuffle and short.

Así que la función *reduce* va a recibir como entrada cada una de estas líneas, por cada key una lista de value. Por cada línea el *reduce* emitirá la suma de las puntuaciones de cada usuario:

```
(Juan, 48)
(Paula, 76)
(Pablo, 65)
```

Tabla 10: Ejemplo MapReduce, resultado proceso reduce.

## 4.3. HADOOP 2.0



Versión de Hadoop con nuevas características para tratar de mejorar el rendimiento del sistema. En la siguiente imagen se puede apreciar la evolución:

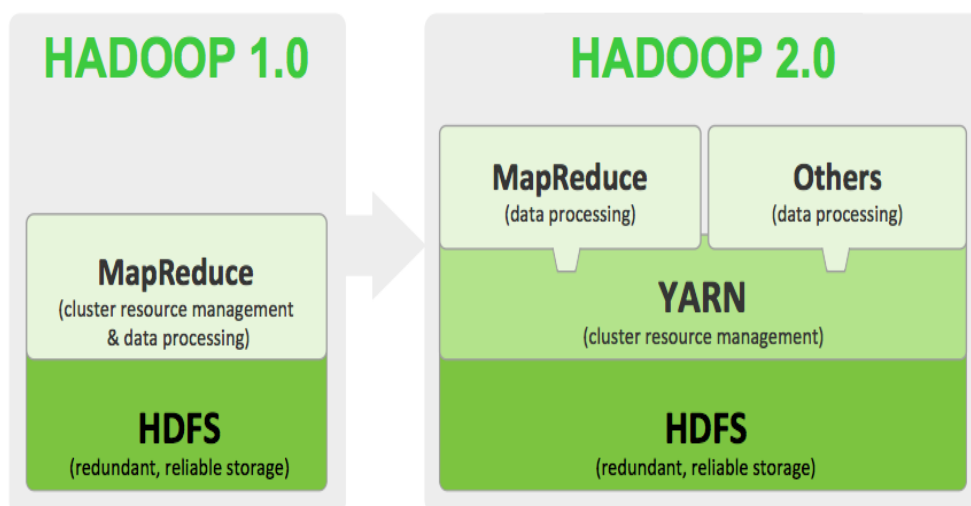


Figura 25: Evolución Hadoop 1.0 a Hadoop 2.0 fuente Big Bang Data Fundación Telefónica.

### 4.3.1. HDFS 2.0: QUORUM JOURNAL MANAGER

La versión HDFS 1.0 presentaba problemas de integridad ante cualquier fallo en el NameNode, en la nueva versión con el fin de eliminar ese problema **se configura un NameNode que está en modo standby a la espera de un fallo en el NameNode activo.**



Para mantener la coherencia de los datos entre los dos NameNodes se usan los JournalNodes, cuya función es la de registrar las operaciones que el NameNode activo va realizando. Este conjunto de JournalNodes se llama Quorum Journal Manager. El Standby NameNode va leyendo el estado del sistema gracias JournalNodes de manera que cuando se produce un fallo cambia a NameNode activo.

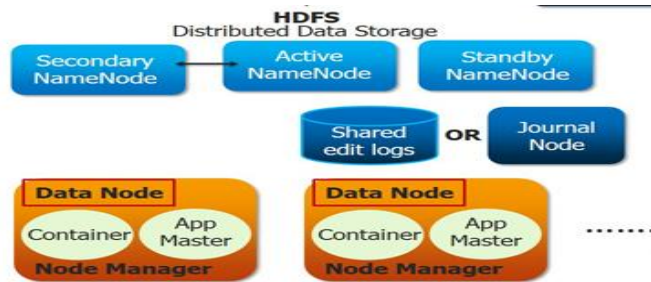


Figura 26: Hadoop 2.0 Distributed Data Storage fuente Big Bang Data Fundación Telefónica.

### 4.3.2 MAPREDUCE 2.0 ARQUITECTURA YARN

YARN (Yet Another Resource Negotiator) es una de las características clave de la segunda generación de la versión Hadoop 2.

Divide las dos funciones principales del NameNode, la gestión de recursos y la planificación y monitorización de las tareas o ejecuciones.

YARN combina un administrador central de recursos con los agentes que monitorean las operaciones de procesamiento de nodos individuales del clúster.

#### Arquitectura de YARN

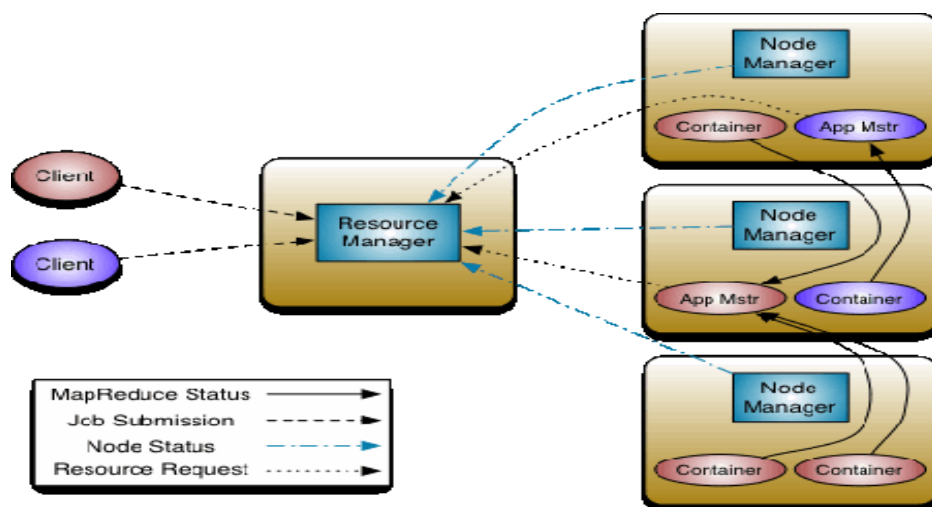


Figura 27: Arquitectura YARN fuente Big Bang Data Fundación Telefónica.

ResourceManager (RM): Elemento global, se encarga de repartir y gestionar los recursos entre todas las aplicaciones del sistema

ApplicationMaster (AM): solicita recursos al Scheduler (responsable de asignar los recursos a cada aplicación) del ResourceManager.

NodeManager: es un agente que se ejecuta en cada máquina y es el responsable de los Containers. Se encarga de gestionar los recursos asignados al Container y reportar su estado al Scheduler. Los Container van reportando su estado al ApplicationMaster, que se encarga de coordinar los procesos.



# 5. DESARROLLO HADOOP EN REDES SOCIALES

---

## HADOOP EN REDES SOCIALES

Existe gran variedad de bibliotecas asociadas a Hadoop. Su objetivo es crear soluciones para cada uno de los escenarios de análisis de datos que hemos visto. Vamos a seleccionar entre todas las opciones la combinación de soluciones óptima para el análisis de redes sociales.

### 5.1 Adquisición de datos y grabación: Flume

En el tratamiento de datos estructurados Sqoop es una de las herramientas más conocidas. Pero la tendencia es hacia el tratamiento de datos no estructurados de las redes sociales, como pueden ser los datos de acceso de logs.



Apache Flume es un servicio que agrega grandes cantidades de datos en **especial log y descarga de datos de redes sociales**. Su arquitectura se basa en la transmisión por streaming. Apache Flume presenta:

- Agentes: procesos encargados de la transmisión de datos.
- Source: encargado de recibir los datos, convertirlos en eventos y escribirlos en el channel. Twitter Source en la imagen.

- Channel: es un almacenamiento pasivo que recibe los eventos del source y espera a que el sink lo consuma, es decir que los sources añaden eventos mientras que los sinks los retiran.
- Sink: se encarga de leer los eventos del Channel y dependiendo de su naturaleza enviarlos a diferentes espacios.

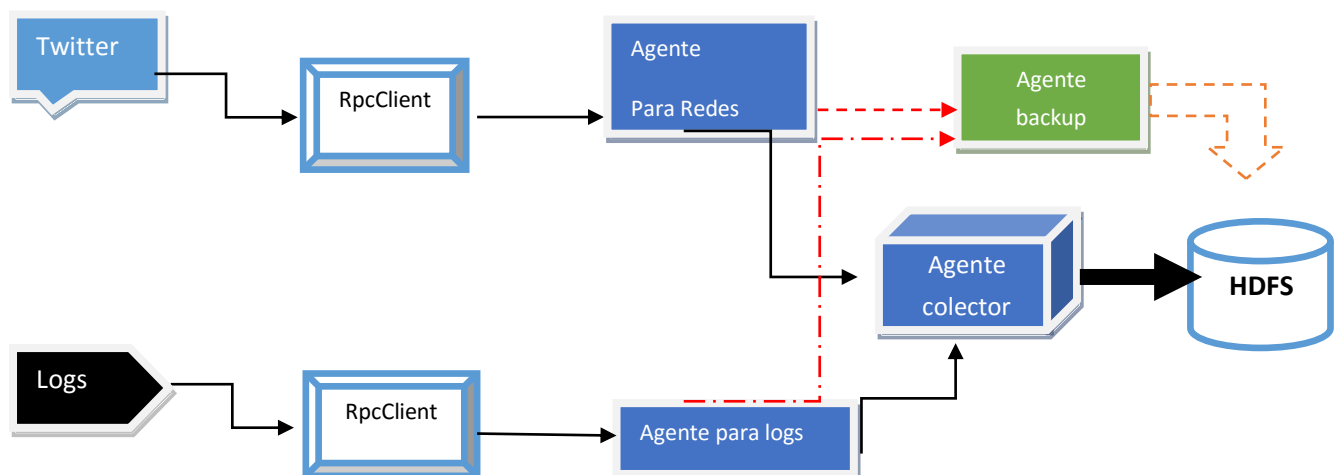
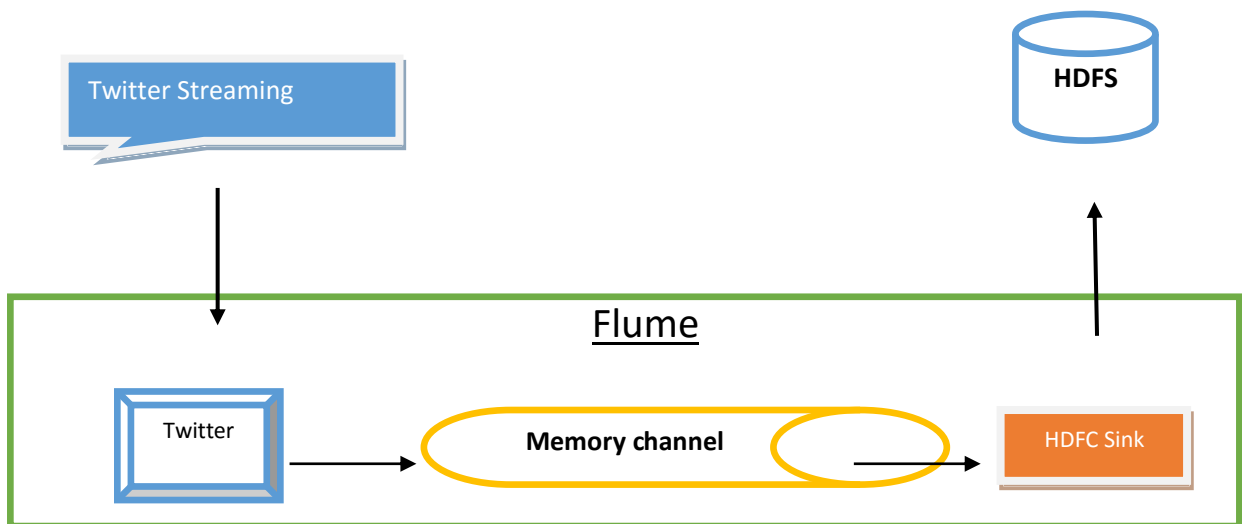


Figura 28: Arquitectura Flume que descarga información de las redes sociales Twitter además de leer ficheros de log.

Flume ofrece una utilidad para facilitar la captura y generación de eventos llamado cliente (RpcClient en la imagen) que trabaja en el origen de datos y que se encarga de recolectar los datos para enviarlo al agente.

En este caso Tenemos dos fuentes de datos, las publicaciones de la red social Twitter y otra fuente de información que son los Logs de usuarios. Se atribuye un agente para cada una de las fuentes dado que en situaciones como la cuenta de la UAM el volumen de publicaciones saturaría el sistema. El siguiente agente “colector” permite unificar la información de ambas fuentes en un solo sistema para que sean enviados al HDFS y no sea necesario duplicar el sistema.

A su vez en color verde se marca un agente de backup por si se produce un fallo de disponibilidad en el agente colector, esté agente solo recibe información en caso de fallo del principal.

## 5.2 Extracción e integración de la información: AVRO

En los primeros capítulos hemos visto que uno de los problemas de Big Data es la diversidad de fuentes y por tanto la disparidad entre tipos de datos a utilizar. Para evitar desarrollar un sistema para cada tipo de dato se hace uso de las librerías que permiten hacer un filtrado previo de los datos dando la estructura necesaria para que sirvan de entrada a **Hadoop**. Alguno de los distintos formatos de la información:



- Texto plano: Arrays de listas de elementos separados por comas.
- Json: Esquema flexible pero poco eficiente en el consumo de espacio, cada campo lleva un nombre asociado.
- Binario: Más eficiente en espacio pero menos flexible, cualquier cambio en formato imposibilita la lectura.

Avro se trata de un sistema de serialización (formato binario) de datos que permite codificar los datos que va a manejar **Hadoop**.

Apache Avro™ es un sistema que ofrece:

- Estructuras de datos complejas.
- Un formato compacto, rápido, binario de datos.
- Un archivo contenedor, para almacenar datos persistentes.
- Llamada a procedimiento remoto (RPC).
- Fácil integración con lenguajes dinámicos.

Las librerías de Avro son de uso sencillo ya que no requieren de generación de código ni en escritura ni en lectura de datos o en comunicación RPC (Remote Procedure Call), una técnica para la comunicación entre procesos en uno o más ordenadores conectada a una red.

Al serializar datos Avro crea un esquema JSON que lo acompañará durante todo el proceso. Avro utiliza JSON, un tipo de codificación muy común y utilizada que se basa en:

- Una colección de pares nombre / valor. En varios lenguajes, esto se crea como object, record, struct, dictionary, hash table, keyed list, or associative array
- Una lista ordenada de valores. En la mayoría de lenguajes, esto se realiza como una matriz, vector, lista o secuencia

El tener el esquema de datos junto con la información serializada permite a la aplicación decidir que realizar en situaciones en las que la aplicación espera recibir otro tipo de datos o esquema distinto al obtenido.

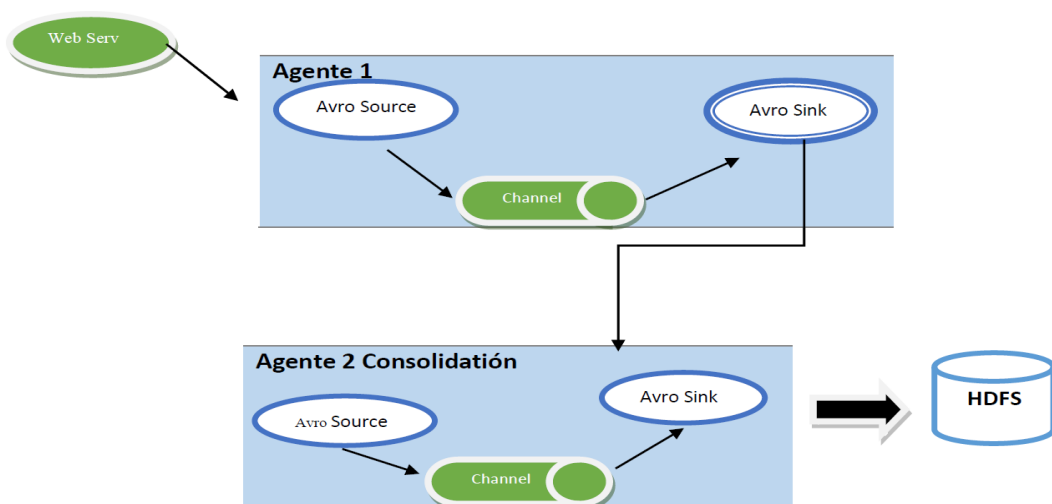


Figura 29: Integración de Avro en la arquitectura anterior.

## 5.3 Procesamiento de peticiones y almacenamiento de datos: CASSANDRA



**Cassandra** es una base de datos **NoSQL** basada en el modelo de almacenamiento de «clave-valor», que empezó como proyecto de *Facebook* en 2009. Es una de las bases de datos NoSQL más relevantes a nivel mundial: Netflix, eBay y Twitter la usan a diario. La arquitectura de Cassandra está basada en una serie de nodos que se comunican con un protocolo *P2P* donde todos los nodos tienen la misma información.

Una red peer-to-peer (P2P) funciona sin clientes ni servidores fijos, sino una serie de nodos que se comportan como iguales entre sí. Los nodos necesitan de una continua sincronía cuando se produce una escritura para mantener la coherencia.

Cuando usamos Cassandra no debemos especificar los campos, lo que permite añadir o eliminar campos.

Cuando un usuario realiza una operación de lectura o escritura sobre uno de los nodos, el nodo pasa a ser el coordinador. La función del coordinador es la de decidir cuáles son los nodos que deberán contestar la petición del usuario.

La información en las bases de datos relacionales, se almacenan en forma de filas, en Cassandra la información se almacena en columnas con pares key-value.

El modelo de datos de Cassandra consiste en particionar las filas, que son reorganizadas en tablas. Las claves primarias de cada tabla tienen un primer componente que es la clave de partición. Dentro de una partición, las filas son agrupadas por las columnas restantes de la clave. Las demás columnas pueden ser indexadas por separado de la clave primaria.

**Column.** Es la unidad básica en Cassandra por tanto se asemeja a los campos de las bases relacionales. Están formadas por los siguientes campos:

- Nombre: Es el key, array de bytes.
- Valor: Array de bytes.
- Timestamp: Guarda la última vez que se accedió a la columna.



Las columnas se pueden agrupar en los siguientes tipos:

**SuperColumn:** Es una columna compuesta por varias columnas y sus nombres.

Las columnas que la forman están ordenadas, y el número de columnas que se puede definir es ilimitada. Las Super columns, a diferencia de las columns, no tienen un timestamp definido.

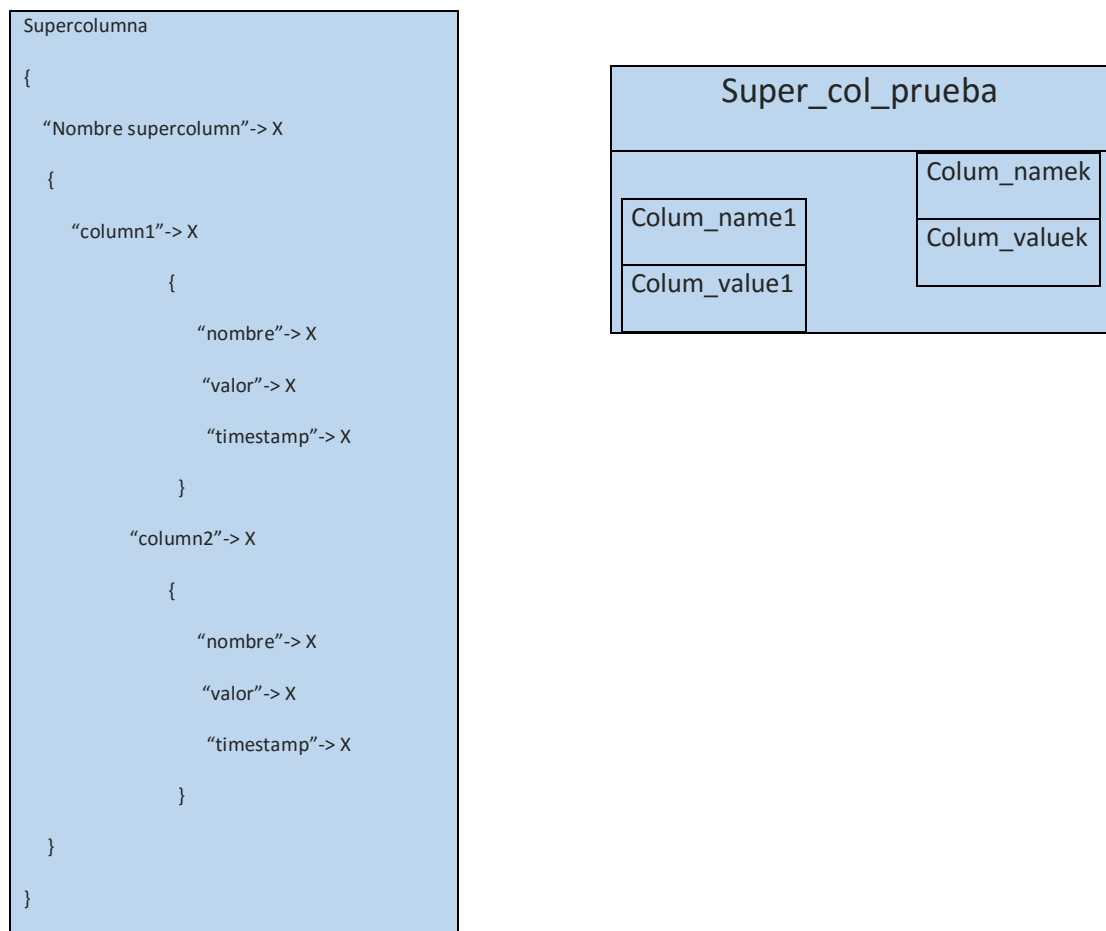


Figura 30: Ejemplo de super columna en Cassandra.

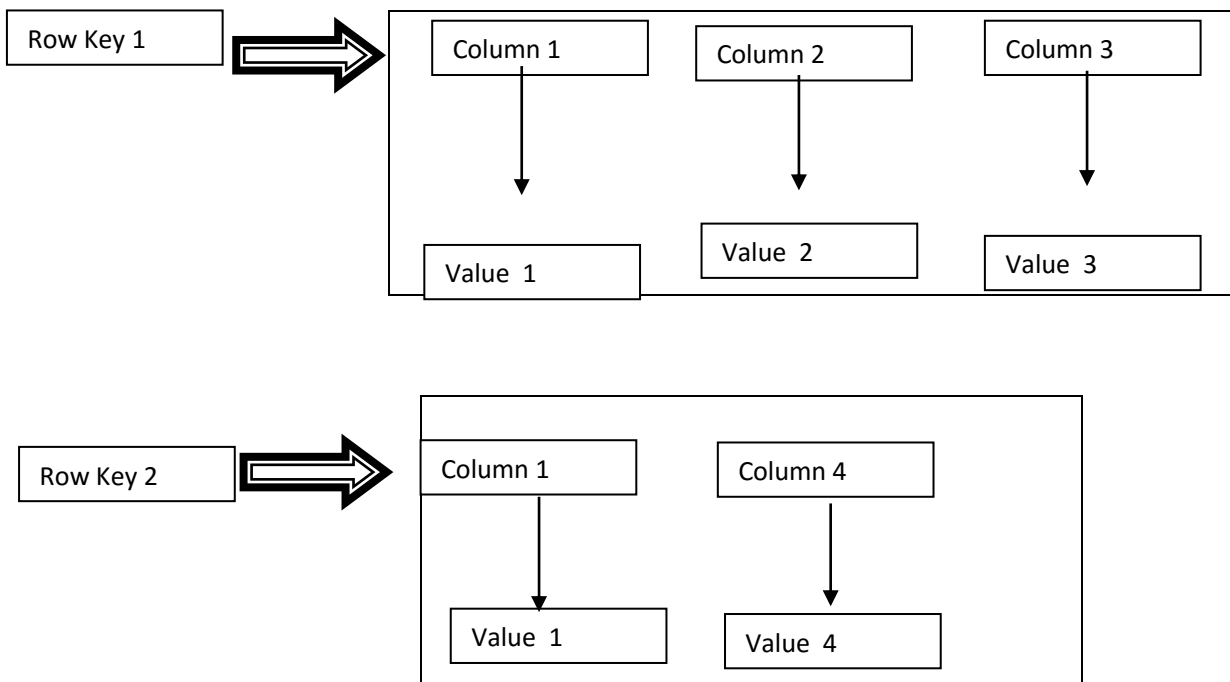
**Family Colum:** Contenedor de columnas más o menos análogo a una tabla en el modelo relacional. Contienen una lista ordenada de columnas. Cada columnfamily se guarda en un fichero ordenado por clave de fila. El contenido consta de un conjunto de filas con un conjunto de columnas similares pero no idénticas. Pueden ser de dos tipos:

- SupercolumnFamily: Mapa ordenado de nombres de súper columnas a mapas de nombres de columnas a valores de columnas.

Row Key 1	Super Column key 1			Super Column key 2			...
	Subcolumn key 1	Subcolumn key 2	...	Subcolumn key 3	Subcolumn key 4	...	
	Column Value 1	Column Value 2	...	Column Value 3	Column Value 4	...	

Tabla 11: Ejemplo de SupercolumnFamily.

- SimplecolumnFamily: es un mapa de nombres de columna ordenadas a valores de columna.



Keyspace. Es el contenedor para las column family. Es más o menos análogo a una base de datos en un modelo relacional, usado en Cassandra para separar aplicaciones. Un keyspace es una colección ordenada de columns family.

Clúster. Conjunto de máquinas que dan soporte a Cassandra y son vistas por los clientes como una única máquina.

## 5.4 CONSULTAS DE DATOS: HIVE



Hive es un almacén de datos creado para Hadoop que permite crear tablas, insertar datos y realizar consultas en un lenguaje similar a como se podría realizar con SQL denominado HiveQL. A través de HiveQL podemos lanzar consultas que serán traducidas a trabajos MapReduce.

Para trabajar con los datos debemos crearnos un esquema, definir la tabla con sus columnas y tipo de dato de cada una de ellas, hacer la carga de los datos inicial y posteriormente realizar las consultas para recoger lo que nos interese.

### Apache Hive Architecture

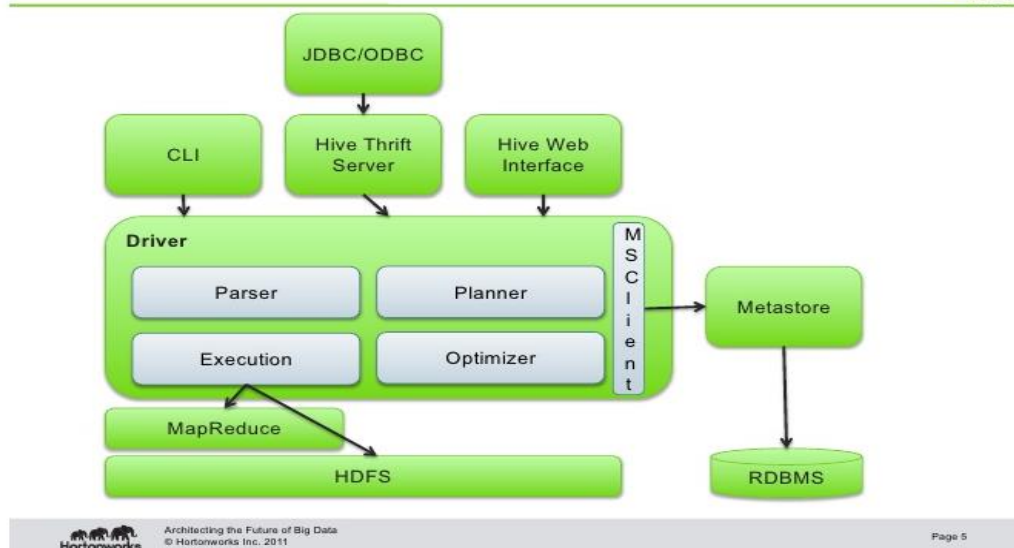
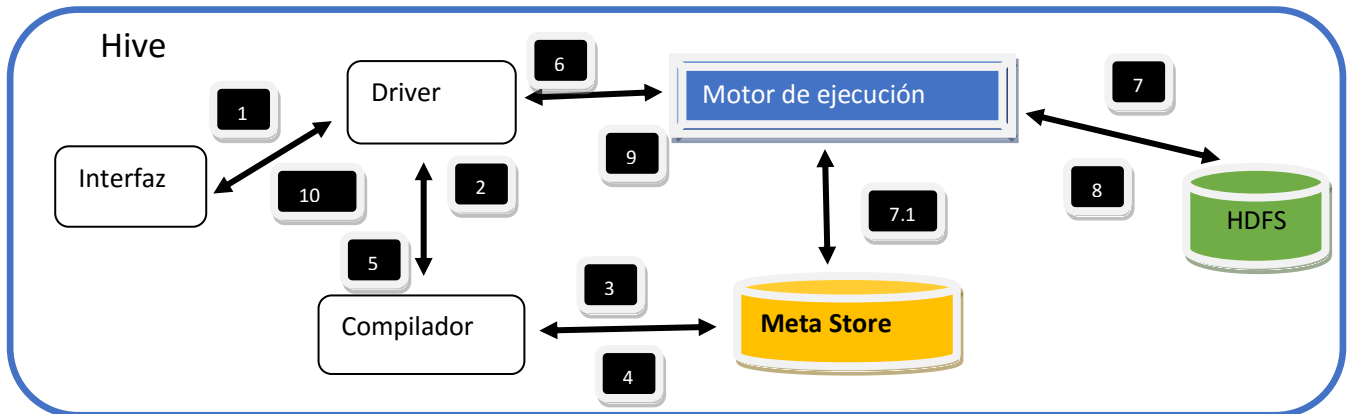


Figura 32: Arquitectura Hive fuente Hortonworks.

## Hive de trabajo

El siguiente diagrama muestra el flujo de trabajo entre Hive y Hadoop.



Funcionamiento	
1	<b>Ejecutar consulta:</b> interfaz de usuario que permite realizar la consulta y se la envía al driver para que se ejecute.
2	<b>Obtener Plan:</b>  Controlador: Recibe las consultas y se encarga de implementar las sesiones.  Compilador: parsea la consulta y realiza análisis semánticos y otras comprobaciones de lenguaje para generar un plan de ejecución con la ayuda del metastore.
3	<b>Obtener metadatos:</b> metastore  Almacena toda la información de la estructura que mantienen los datos dentro de Hive -es decir, tiene el esquema de las bases de datos, tablas, particiones, etc.-.
4	<b>Enviar Metadatos:</b> Metastore envía los metadatos, como una respuesta para el compilador.
5	<b>Enviar Plan</b>  El compilador comprueba el requisito y vuelve el plan al driver. Hasta aquí, el análisis y elaboración de una consulta es completa.
6	<b>Ejecutar Plan:</b> El controlador envía el plan a ejecutar el motor de ejecución.
7	<b>Ejecutar trabajo</b> Internamente, el proceso de ejecución se trata de un trabajo de MapReduce.
7.1	<b>Metadatos:</b> Mientras tanto en la ejecución, el motor de ejecución puede ejecutar las operaciones de metadatos con Metastore.
8	<b>Buscar resultados:</b> El motor de ejecución recibe los resultados de nodos de datos.
9	<b>Enviar Resultados:</b> El motor de ejecución envía los valores resultantes para el driver.
10	<b>Enviar Resultados:</b> El controlador envía los resultados a Hive Interfaces.

## 5.6 Interpretación de datos.: Mahout



Los algoritmos de clasificación son algoritmos de aprendizaje, en los que a partir de ejemplos se calculan patrones para obtener un modelo.

**Apache Mahout**, es un proyecto de software de la Fundación Apache para crear una biblioteca de aprendizaje automático (e inteligencia colectiva) escalable usando Hadoop.

Desde "*Introducing Apache Mahout*" se publicó por primera vez en developerWorks la comunidad Mahout — y el código base y capacidades del proyecto — han crecido significativamente. Mahout también ha visto una acogida significativa por compañías grandes y pequeñas en todo el mundo.

Tamaño del sistema	Mahout o sistemas tradicionales.
<100 K	<b>Pocos datos, Mahout resulta lento en los procesos.</b>
1M-10M	<b>Mahout.</b>

Tabla 12: Mahout frente a métodos tradicionales por usuario.

Mahout tiene un gran número de algoritmos implementados para trabajar con Machine Learning:

- **Clustering**: para descubrir estructuras en colecciones de datos.

– **Classification:** es una técnica supervisada para descubrir relaciones, por ejemplo en base a los mails spam etiquetar nuevos mails como spam.

– **Collaborative filtering:** técnica para recomendaciones, por ejemplo recomendar otros temas que pudieran gustarles en relación a temas que ya les gustan.

Acepta datos en formato de **userID, itemId y prefValue** (la preferencia de los usuarios por el elemento). Mahout puede realizar entonces análisis de ocurrencias conjuntas. A continuación se muestra un ejemplo:

- **Ocurrencia conjunta:** a Oscar, Gonzalo y David les gusta *El Señor de los Anillos: la Comunidad del Anillo*, *El Señor de los Anillos: las dos torres* y *El señor de los anillos: el retorno del rey*. Mahout entiende el patrón de unión entre las tres películas y que a los usuarios que les gusta alguna de estas películas también les gustan las otras dos.
- **Ocurrencia conjunta:** a Francisco y Javier también les gusta *El hobbit: un viaje inesperado*, *El hobbit: la desolación de Smaug* y *El hobbit: la batalla de los Cinco Ejércitos*. Mahout determina que a los usuarios que les gustan las tres películas anteriores también les gustan estas tres. Mahout entiende que existe otro grupo nuevo que continua con este tipo de patrón y por ello entiende que existe relación de gusto entre ambas trilogías.
- **Recomendación basada en similitud:** como a José le gustan las tres primeras películas, Mahout examina películas que a otros usuarios con preferencias similares les han gustado, pero que José no ha visto (gustado/valorado). En este caso, Mahout recomendaría *la trilogía del Hobbit*.

Entre las funciones cuenta con algunos de los algoritmos más conocidos en el mundo del aprendizaje: los clasificadores Naive Bayes y Random Forest, arboles de decisión o funciones de clustering como k-means o Fuzzy k-means.

## Terminología de los clasificadores

**Modelo:** La salida del algoritmo de entrenamiento. Nos permite prever comportamientos futuros de los individuos bajo estudio o posibles variaciones por ejemplo en el mercado de la bolsa.

**Datos de entrenamiento:** Muestras de entrada etiquetada que se introducen al sistema con el fin de que sea capaz de encontrar patrones y crear una respuesta lineal. Cuanto mayor es el número de muestras de entrenamiento más fiel es el modelo.

**Datos de prueba:** Datos de entrenamiento con el valor de la variable objetivo ocultos para que puedan ser usados para evaluar el modelo. Es un modo de “encaminar” al sistema hacia el patrón que quieres estudiar.

**Entrenamiento** El proceso de aprendizaje que usa los datos de entrenamiento para producir un modelo. Este modelo puede estimar la variable objetivo dada por las variables predictores como entrada. Por ejemplo en la creación de un reconocedor de palabras tener diferentes personas con distintos tonos diciendo esa misma palabra.

**Ejemplo de entrenamiento:** Una entidad con características que serán usadas como entrada para el algoritmo de aprendizaje. Ser capaces de encontrar los ejemplos de entrenamiento de forma objetiva y en la cantidad necesaria para crear modelos reales.

**Variable:** El valor de una característica o una función de varias características.

**Registro** Un contenedor donde se almacena un ejemplo. Está compuesto de campos. Puede ser por ejemplo una grabación de voz etiquetada.

**Campo** Parte de un registro que contiene el valor de una característica (variable).

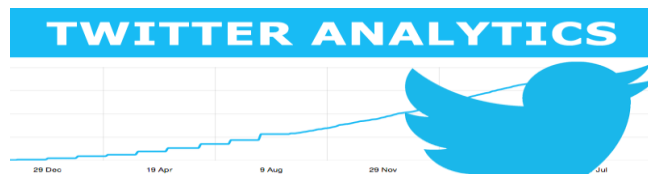
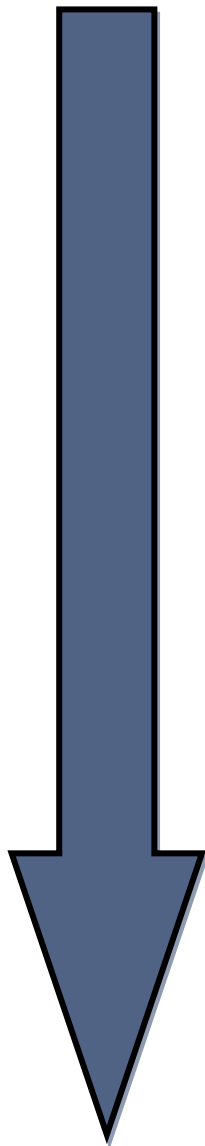


**Variable predictora** Una característica seleccionada para usarla como entrada al modelo de clasificación. No todas las características necesitan ser usadas. Es el modo de intentar actuar con el sistema viendo solo entrada controlada y salida esperada.

**Variable objetivo** Una característica que el modelo de clasificación intenta estimar.

# PROCESO COMPLETO

---





## 6. BIG DATA EN LAS REDES SOCIALES DE LA UAM

---

### 6.1 ¿Qué relación tienen el Big Data y las redes sociales?



El análisis de grandes conjuntos de datos en la educación universitaria puede proporcionar pautas para su mejora. Aunque las universidades vienen recogiendo datos sobre los estudiantes desde hace mucho tiempo, la mayoría de ellos relacionados con su rendimiento escolar, la novedad es la implementación y desarrollo de plataformas 2.0 para la difusión de material educativo e interaccionar con los estudiantes. Casos típicos y de éxito son los *MOOCs*, *Youtube*, *blogs*, aplicaciones móviles y redes sociales para el aprendizaje y la autoevaluación. Esta metodología ofrece numerosas ventajas, porque permite a los estudiantes acceder a la enseñanza de forma inmediata y proporciona una realimentación inmediata y constante sobre su interés y rendimiento.

Las redes sociales se han popularizado entre los estudiantes. **Facebook, por ejemplo, presenta más de mil millones de usuarios, Twitter 500 millones y LinkedIn 225 millones (IAB 2014).** Así, junto al uso meramente social, como espacio y vía de comunicación, información y entretenimiento; las redes sociales se han convertido en un espacio idóneo para intercambiar información y conocimiento de una forma rápida, sencilla y cómoda. Los docentes aprovechan esta situación y la predisposición de los estudiantes a usar redes sociales para incorporarlas a la enseñanza.

En este escenario se desarrolla el *Social Media Intelligence*, utilizar los datos que ya están en las redes sociales para aumentar la competitividad de una institución o un laboratorio universitario. La idea es conocer que piensan acerca de “nuestra marca” y tener acceso a estas opiniones en tiempo real. Todo esto evitando estudios de mercado costosos.

“*Twitter* se ha convertido en el mayor archivo público de la conducta y pensamiento humano que ha existido nunca. Sus datos se utilizan para todo, desde predecir

epidemias hasta ayudar en la respuesta de los desastres naturales. De hecho, hemos puesto esta información a disposición de los investigadores académicos a través de **Twitter Data Grants** porque creemos que apenas estamos arañando la superficie de todo lo que se puede conocer”, explica *Alex Gibelalde, responsable de Marketing de Twitter España*.

Según el informe de **AIMC**, España cuenta con una población online de 24,8 millones de personas. El 71% de esta población (17,6 millones de usuarios) utiliza activamente las **redes sociales**.

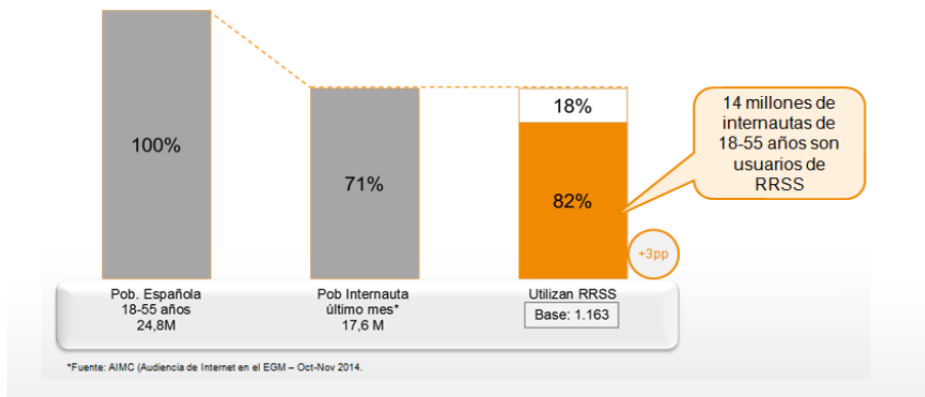


Figura 33: Audiente de Internet y número de usuarios en RRSS de 2014

Un 75% dice que accede a las redes desde su Smartphone y el 28% desde Tablet. Aparece ahora como nuevo integrante el acceso desde SmartTV o desde plataformas de gamesharing como *PlayStation Networks*.

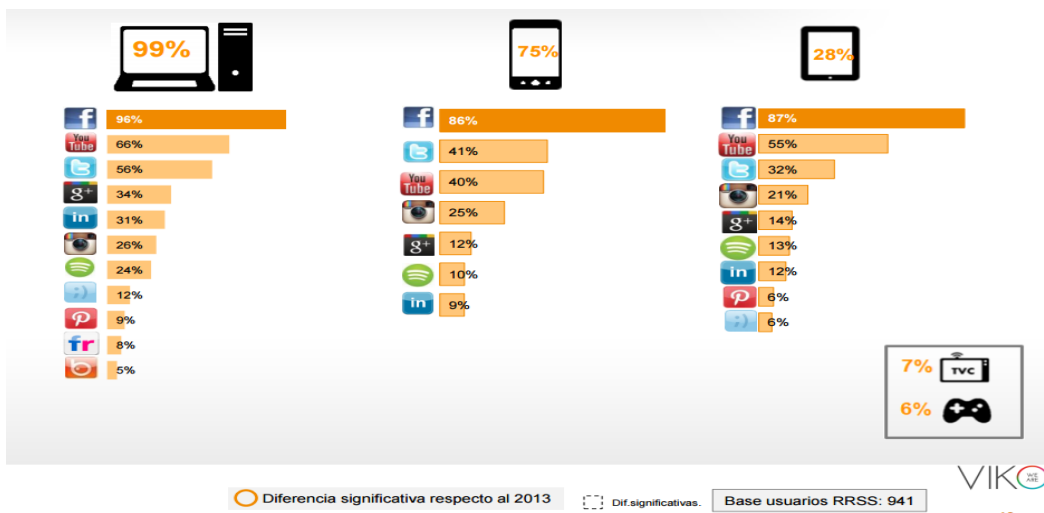


Figura 34: Dispositivos de acceso a las diferentes redes sociales fuente Viko.

Por importancia, las tres redes más usadas en 2015 por los internautas españoles son **Facebook, Youtube y Twitter**. El 88% de los españoles que utilizan Internet tiene cuenta en **Facebook** (frente al 87% en 2013) y el 56% en **Twitter** (54% en 2013). La principal red profesional es **LinkedIn**, con un 32% de usuarios españoles sobre el total.



## Monográfico adolescentes Comparativa adolescentes vs resto población

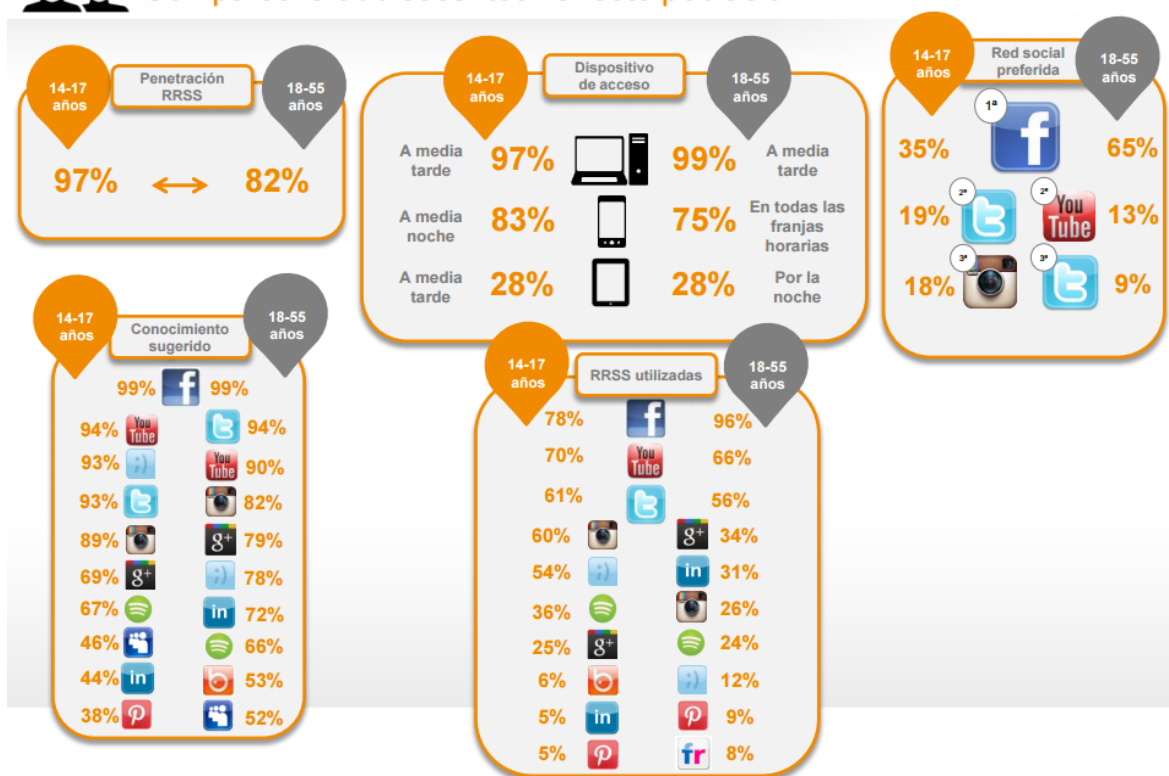


Figura 35: Monográfico adolescentes en redes sociales, fuente IAB.

Las redes sociales por tanto son el escenario perfecto para la actividad centrada en el análisis de las necesidades del estudiante tanto como para ampliar la imagen internacional institucional por su gran proyección. En el siguiente apartado vamos a analizar los datos referentes a los estudiantes de la Universidad Autónoma de Madrid desde horarios de actividad, geografía, intereses profesionales, dispositivos de acceso...

En relación a lo desarrollado previamente se muestra el esquema para obtener la analítica de dichos datos con un breve esquema.

1. Recolección de datos de la red social elegida.
2. Almacenamiento de la información en bases de datos internas o externas. Por el volumen de la información y para evitar problemas de seguridad se ha optado por usar los sistemas de almacenamiento de SocialBro para evitar infracciones legislativas.
3. Dependiendo la red social los datos presentan diferentes formatos. En caso de Twitter la información obtenida tiene formato txt. Posteriormente se pueden aplicar filtros para conocer geolocalización u horarios.
4. Visualización de datos: mediante la exportación de datos a Excel o mediante Google Tables.

## 6.2 big data redes sociales: TWITTER UAM



Creada en el 2006, Twitter permite publicar contenido de máximo 140 caracteres. Twitter se la conoce como una red social de microblogging. Cada día se envían 500 millones de tuits y cuenta con 302 millones de usuarios activos, siendo, por el momento, la tercera red social más popular a nivel global.

Informar acerca de las actividades organizadas por los centros, difundir datos de interés para los alumnos, como la convocatoria de becas, responder de manera ágil a las dudas de estos o contactar con ellos de una forma distendida. Permite al estudiante configurar la información UAM que desea ver.

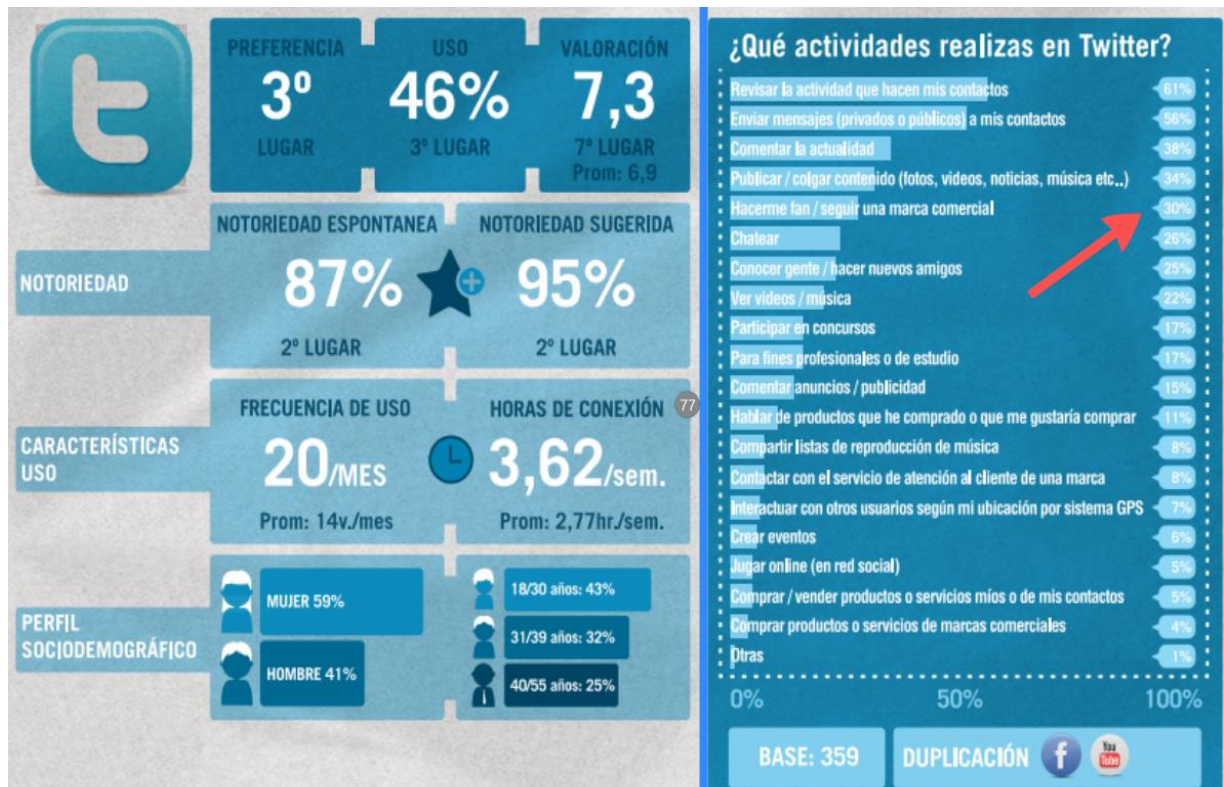


Figura 36: analisis sociodemografico y de uso de la red social Twitter fuente IAB.



El 80% de los usuarios de esta red social accede desde su Smartphone dado que somos más proclives a dedicar nuestro tiempo a Twitter cuando estamos fuera de casa que otros redes sociales (181%) sobre todo por la facilidad para conocer y compartir contenidos.



El uso de Twitter es principalmente profesional y académico (61% de los encuestados (mientras que el 19% lo utiliza de forma personal y el 21% combina ambas facetas indistintamente.

32% de internautas usa Twitter



Figura 37: Estadísticas por sexo e intervalos de edad de la red social Twitter fuente IAB.

Según el estudio **"Nielsen: Respuesta neurológica predice actividad en redes sociales"**, Twitter es una de las variables significativas que influye en los niveles de audiencias televisivas. El estudio establece una correlación entre el incremento de tuits y el aumento de la audiencia para el programa sobre el que se está tuiteando. Es el fenómeno conocido como "Segunda pantalla". **En España en 2013, 4.5 millones de usuarios tuitearon sobre contenido televisivo.**



## 6.2.1 twitter uam: análisis mediante socialbro

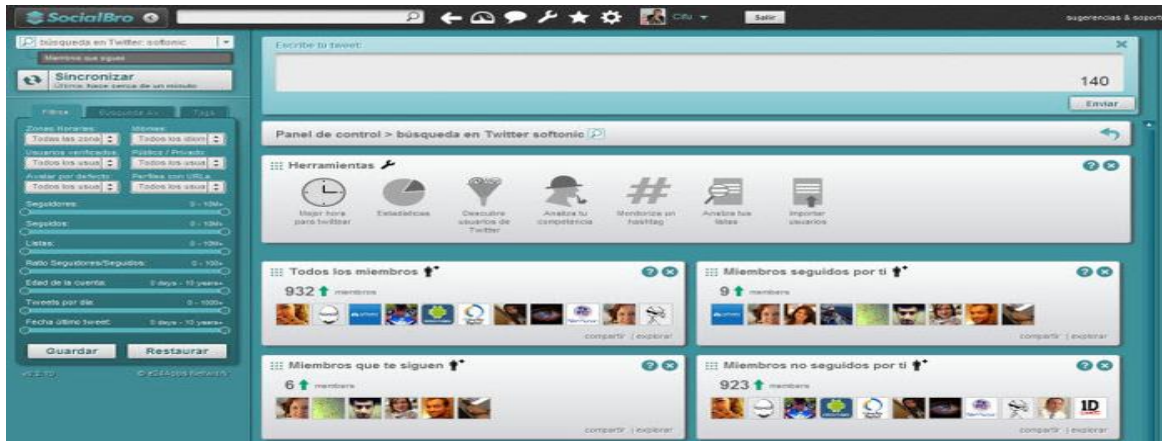


Figura 39: Interfaz de usuario de SocialBro.

Aplicación multiplataforma que corre sobre Adobe AIR. Esta herramienta permite conocer en profundidad los datos de tus seguidores y seguidos: intereses, características demográficas y sus comportamientos. Permite a su vez conocer los horarios de máxima actividad de la comunidad para aumentar nuestra presencia online.

Uno de los últimos estudios ha demostrado que la población de *Twitter* es más activa los domingos y en horarios referentes a las 22:00. Debemos ver para cada contenido la mejor hora y cubrir la franja horaria del fin de semana porque es cuando los alumnos tienen más tiempo para poder ver *Twitter*.

Podemos comparar directamente nuestras estadísticas con 2 competidores, por tanto conocer nuestra posición a nivel regional, nacional e internacional con la comunidad universitaria.

Otra manera de analizar el contenido es por las palabras clave llamadas hashtags que permiten a varios usuarios sumarse en una misma conversación sobre un evento o tendencia. Por ejemplo descubrimos las personas que han tuiteado sobre #UAM. Marcando estos hashtags somos capaces de segmentar la comunidad universitaria.

Twitter	Dato Enero 2015	Variación respecto Diciembre 2014
Follows	+476 (Total 32.282)	+1.5%
Tuits publicados	184	+8.2%
Menciones (@)	1330 (ratio 7.23)	+12.8%
Favoritos	390	+3.2%
RT	589	-8%

Tabla 13: Comparativa de datos obtenidos Twitter Enero 2015vs Diciembre 2014 vía Socialbro.

### KREDde score evolution de la UAM:

- 1) Influencia: Se basa en una escala de 1000 puntos, medidos teniendo en cuenta la frecuencia de retweets, menciones y follows en Twitter. Influencia (859), un 0,1 por ciento más que en diciembre 2014.
- 2) Alcance (Outreach). Refleja la generosidad a la hora de interactuar con los demás y difundir su mensaje. Actualmente, el nivel máximo de Outreach alcanzado es 12. ; Alcance (7), sin cambios.

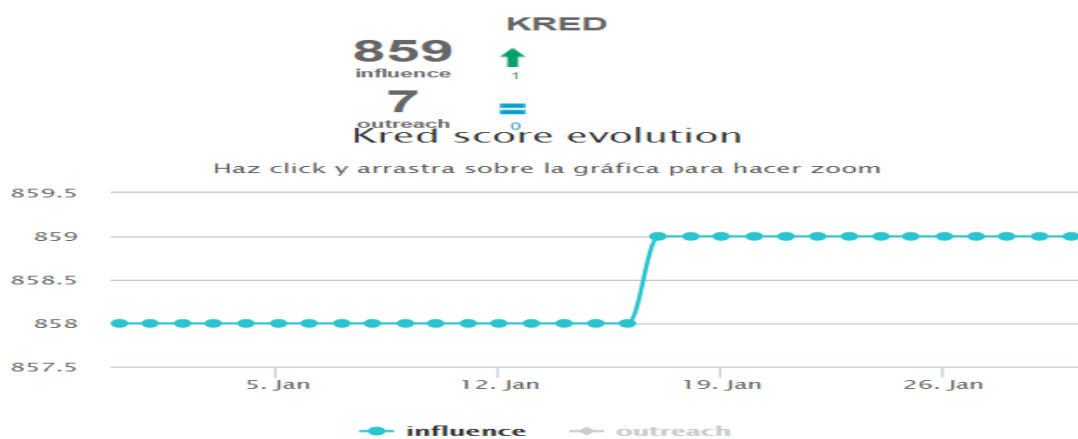


Tabla 14: Kred score evolution durante el mes de Enero 2015.

## RATIOS DE ENGAGEMENT

(RT “amplificación”, FAV “aplauzo” y menciones “conversación”): ER= (personas hablando de esto / número total de fans) x100.

Tabla 15: Ratios de engagement de Twitter de la Universidad Autónoma de Madrid.

### RATIOS DE ENGAGEMENT



La importancia de estos datos no solo tiene que ser cuantificable y en búsqueda de aumentar estadísticas sino que es realmente importante obtener como en la imagen adjunta los datos de las publicaciones de mayor éxito. Aparece reflejado a su vez el “potencial” que es el número de personas que han “visto” el tuit y han sido por tanto target de nuestra publicación.

**Tabla 16: Evolución de seguidores durante Enero 2015 y previsión Febrero de Twitter de la Universidad Autónoma de Madrid.**



El análisis de los seguidores obtenidos durante el mes de enero crece de forma prácticamente lineal como se puede apreciar en la gráfica. Se ha realizado una estimación para el siguiente mes pero puede no ser muy acertada dado que en época de exámenes los estudiantes suelen proyectar más atención a las redes sociales en búsqueda de información y avisos relevantes.

También se adjunta una gráfica de “me influyen” es decir personas con las que hemos interactuado o compartido sus publicaciones.

## ESTADÍSTICAS DE CONTENIDO

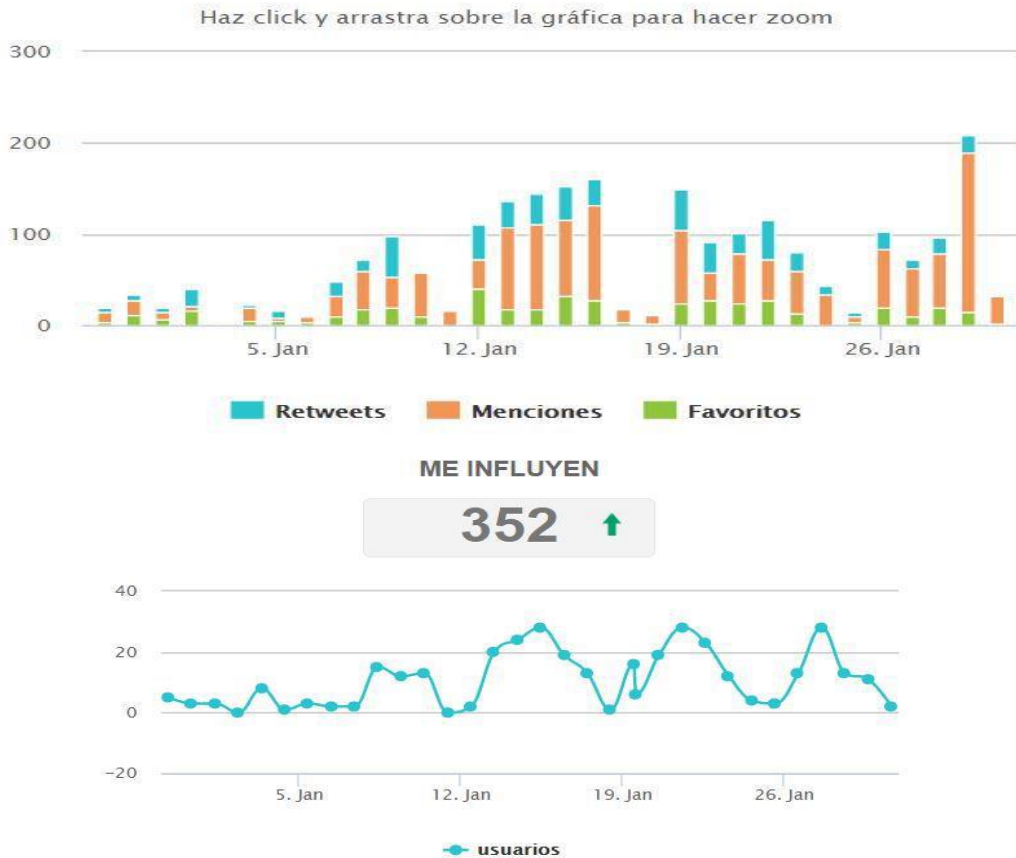


Tabla 17: Gráfico detallado de la interacción con los usuarios con nuestras publicaciones.

En la gráfica adjunta se añade un gráfico desglosado del éxito e interacción de las publicaciones en intervalos de semanas. Nos permite por tanto de una manera rápida ver qué días las personas están más comunicativas (nos mencionan). Podemos ver por tanto un análisis de una semana los días más receptivos para publicar recursos o preparar nuestros recursos para los días de mayor actividad.

Previamente se ha citado la importancia de realizar la búsqueda de aquellas personas que están “hablando” de un tema determinado. Twitter determina sus tendencias en función del número de personas escribiendo sobre una misma tendencia haciendo escalar siempre aquellas tendencias más novedosas, esa es la razón por la cual tendencias como Justin Bieber no son siempre trending topic. Es importante conocer aquellas palabras más repetidas en nuestra comunidad universitaria para conocer sus gustos.

Por ejemplo si una muestra grande de usuarios habla de #derecho justo antes de los exámenes de selectividad podemos suponer que ese año tendrá una gran demanda de inscripción esa titulación. Podemos por tanto conocer cuántas personas hablan de #UAM o lo tienen incluso en su biografía para invitarles a conocer nuestras redes sociales por ejemplo. Es el mejor método para segmentar comunidades en internet o conocer incluso cuantas personas están hablando de nuestra competencia.

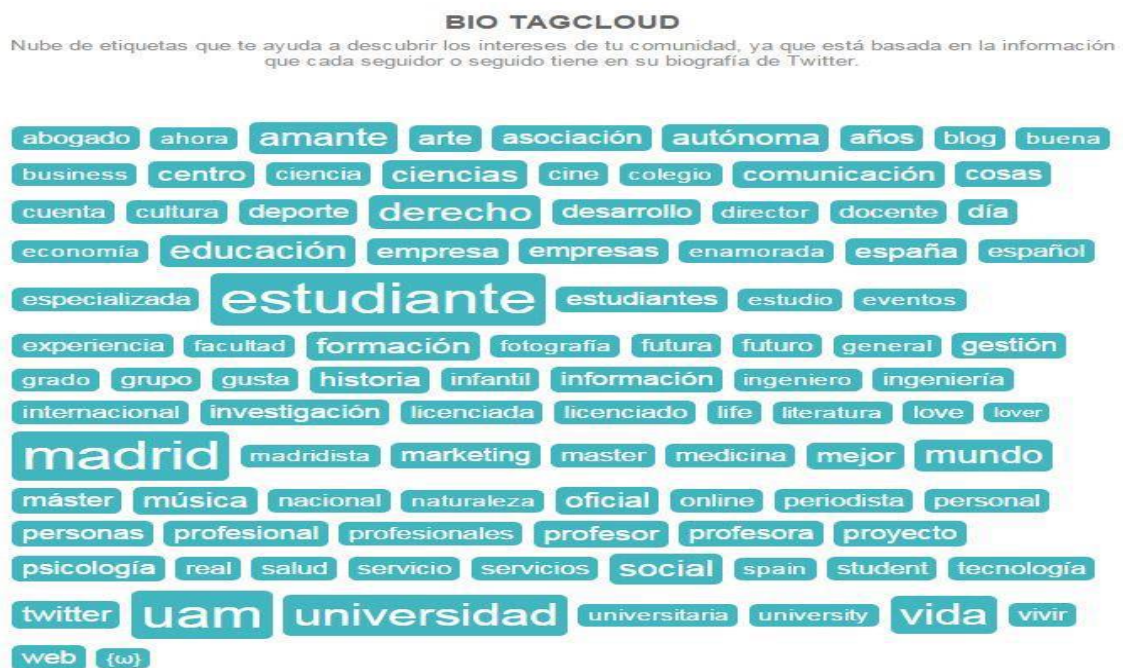


Tabla 18: Mapa de palabras más utilizadas por la comunidad de seguidores de la UAM.

Localizadas las personas de la comunidad el siguiente aspecto importante es conocer la localización de nuestra comunidad. Es un modo eficiente de conocer si nuestra estrategia internacional está siendo exitosa y los puntos débiles donde debemos reforzar nuestra presencia. En la Universidad Autónoma de Madrid la política internacional de captación de estudiantes en Sudamérica es muy exitosa pero también podemos analizar esa política a nivel nacional. La localización del público es fundamental tanto para utilizar un idioma determinado (en el caso de la UAM hispanohablantes en su mayoría) y conocer los horarios más determinantes por la

## UBICACIONES

Tagcloud de ubicaciones de tu comunidad.



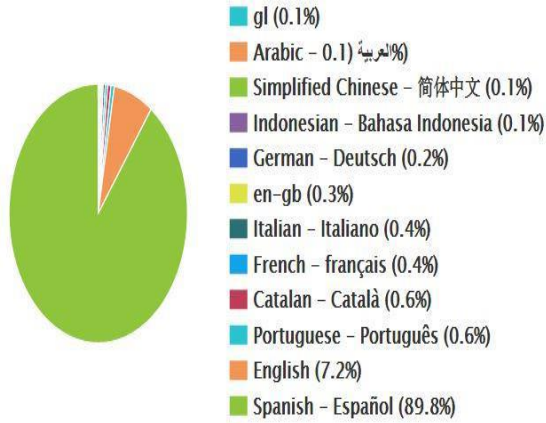
diferencia de husos.

Tabla 19: Tagcloud de las ubicaciones de la comunidad de usuarios de la Universidad Autónoma de Madrid.



### IDIOMAS

En esta gráfica puedes ver los principales idiomas de tu comunidad, útil a la hora de decidir en qué lenguas publicar tus tweets.



### ZONAS HORARIAS

Esta gráfica te enseña las zonas horarias en las que está tu comunidad. Esta información puede ser útil a la hora de crear un informe de la Mejor Hora para Twitter.

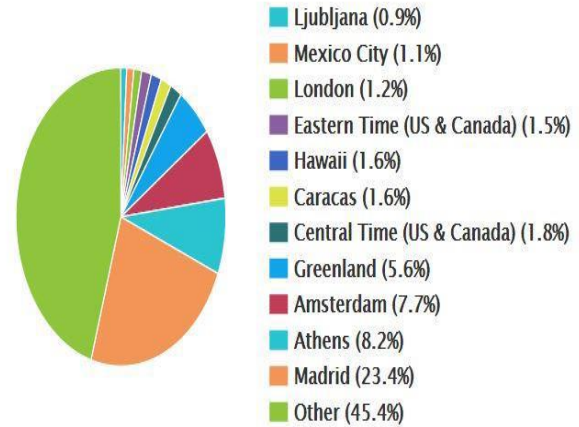
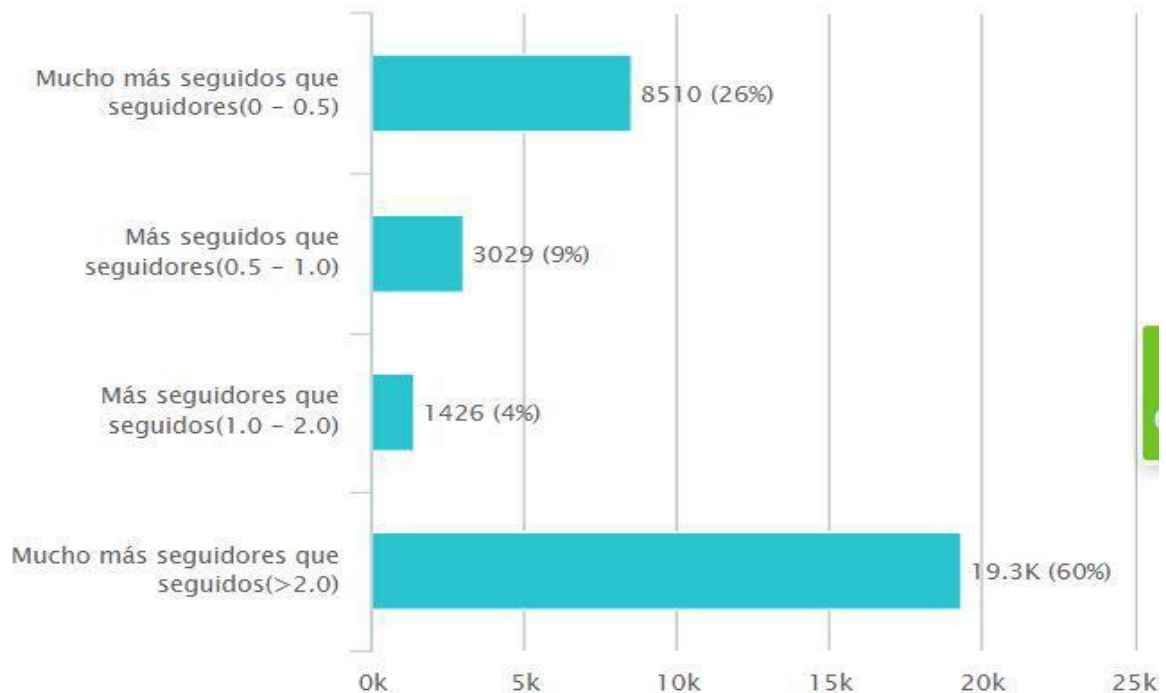


Tabla 20: Estadísticas de los idiomas de nuestra comunidad y las zonas horarias de mayor actividad de la UAM:

### RATIO SEGUIDORES/SEGUIDOS

Esta gráfica te da una idea del nivel de influencia de los miembros de tu comunidad. Las personas que tienen un ratio por encima de 2 tienen mucho más seguidores que seguidos, lo que significa que son más influyentes que las demás.

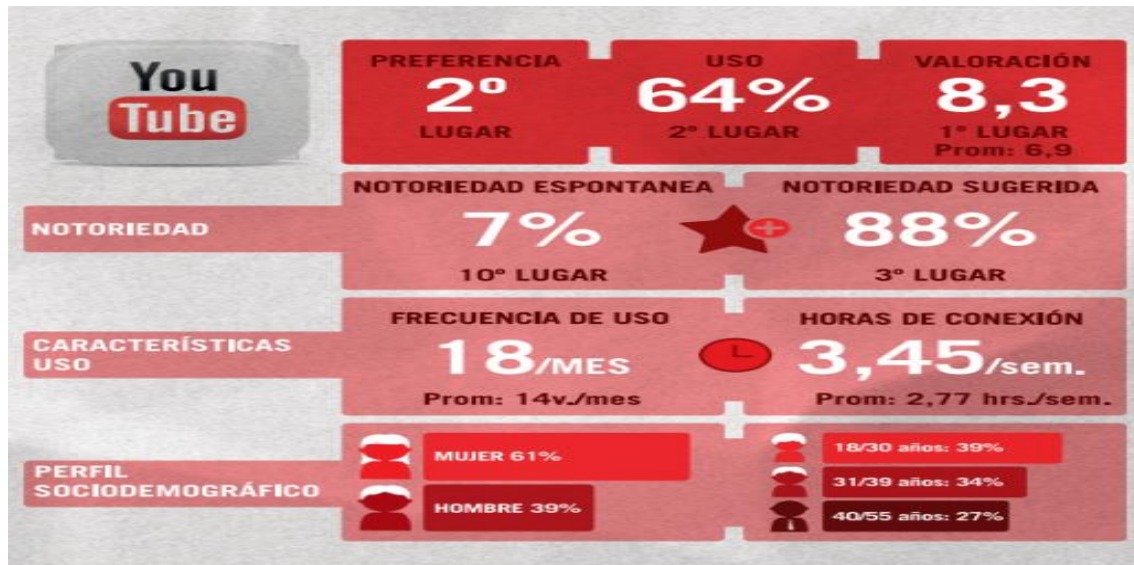




### 6.3 big data y redes sociales: Youtube y los MOOCs.

YouTube es un sitio web caracterizado por la posibilidad de subir y reproducir videos. Fue fundado en el año 2005 y al día de la fecha ha mantenido un crecimiento constante, generando un canal de socialización a partir del comentario de los distintos videos que alberga.

Figura 40: analisis sociodemografico y de uso de la red social Youtube fuente IAB.



Eric Grimson, científico canadiense experto en informática y asesor del Instituto de Tecnología de Massachusetts ha impulsado los denominados cursos MOOC (**Masive Open Online Course**). Grimson habla de que la tecnología va a revolucionar en un futuro próximo la educación. Desde la forma de acercarse el profesor al alumno hasta la programación de los propios títulos oficiales para introducir contenidos *online*.

• Mayo 2013



El 93% de los niños entre 8 y 11 años usa Youtube. “Los MOOCs pueden responder preguntas y a gran variedad de cuestiones y tienen *unfeedback* al instante.”

“Eric Grimson”

• Proyección 2017



Fuente: Hollybyte (ComScore).

En marzo de 2014 nuestra Universidad se ha incorporado al consorcio edX para difundir a través de esta plataforma sus cursos MOOC. EdX fue fundada por el *Instituto Tecnológico de Massachusetts y la Universidad de Harvard* en 2012 para alojar cursos gratuitos online de las más diversas disciplinas.

Según *el MIT Technology Review*, la educación online ha demostrado ser en algunos casos más exitosa que las clases presenciales tradicionales. Gracias a los informes sobre el progreso de los alumnos, los datos sobre el tiempo de dedicación y el éxito de las sesiones de coaching, el número de estudiantes que consigue aprobar se sitúa en el 75%.

En **“La revolución de los datos masivos”** Kenneth Cukier describe cómo el universo educativo puede beneficiarse de los MOOCs:

“Lo que ha sucedido en Estados Unidos, con los cursos en línea, llamados MOOCs, cursos masivos abiertos en línea, es que los profesores pueden ver cuando los estudiantes están viendo sus cursos, y cuando se detienen, cuando releen una lección. Y un profesor de Standford se dio cuenta de que hacia la lección siete u ocho todos los estudiantes regresaron a la lección número tres. Esa lección era una clase de repaso de matemáticas y mostraba que a medida que los estudiantes avanzaban más en el curso, estaban menos seguros de sus bases en matemáticas. Normalmente, un profesor no tiene porqué saber que la clase se está quedando atrás... pero de repente, el profesor podía ver esto en los datos y podía aprender dos cosas. En primer lugar, que debía preparar mejor a sus alumnos. Y en segundo, que debía insistir más en esa dificultad en particular ya que los estudiantes se detenían y regresaban a esa lección.”

**YouTubeEDU** con 11 millones de suscriptores es un canal creado para universidades donde los que alumnos y profesores pueden compartir sus clases e investigaciones, seminarios y cursos o, mostrar cómo es la vida en el campus. Universidades de prestigio internacional como *Cambridge, Harvard o Columbia* son sólo un ejemplo de los centros educativos que están presentes en *YouTubeEDU*.





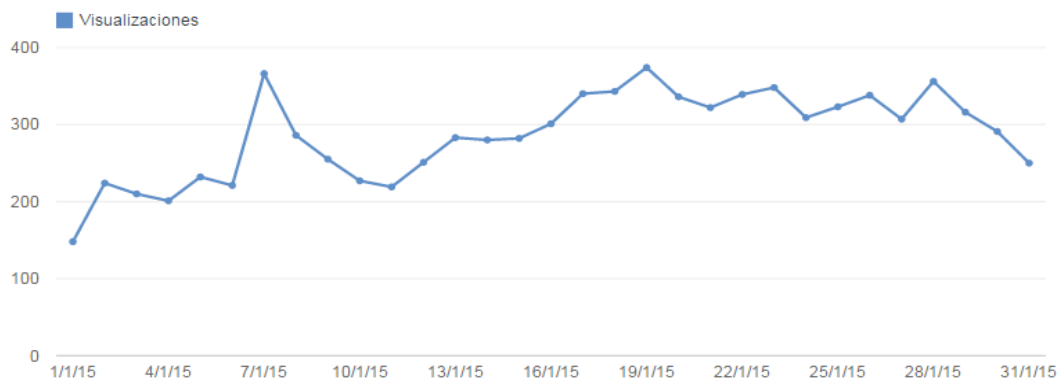
### 6.3.1 Youtube Analytics.

Con YouTube Analytics se puede realizar la analítica del canal y los vídeos mediante métricas e informes en tiempo real.

En primer lugar se ofrecen los datos generales del canal. Desde el número de reproducciones de todos los videos subidos, el número de suscriptores y los videos que han sido de mayor éxito e impacto.

Tabla 21: Estadísticas de visualizaciones de los videos de Enero 2015.

UAM		Datos Enero 2015		Videos más vistos		Reproducciones	
Videos subidos		3 (Total 711)		Cómo citar la bibliografía: el estilo de Vancouver		1185	
Reproducciones		+12549 (Total:709.744)		La UAM es Campus de Excelencia		544	
Suscripciones		+21 (Total:1121)		Video Institucional de la UAM		346	



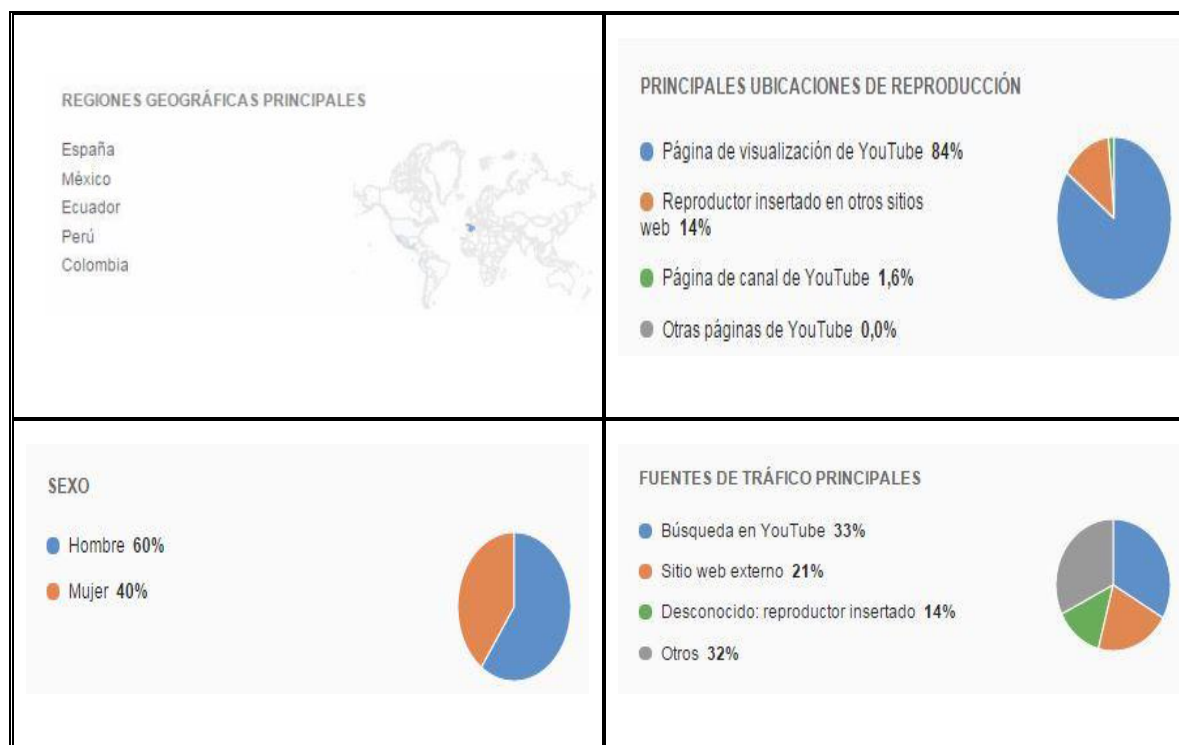


Tabla 22: Análisis sociodemográfico de la comunidad de usuarios de Youtube de la UAM.

Youtube permite segmentar la comunidad de universitarios de nuestro canal. Segmentamos primero por ubicación geográfica. En relación con el posicionamiento web debemos conocer los dispositivos de reproducción y acceso de los usuarios para crear contenidos responsivos (adaptados a Smartphone). Es importante conocer cuáles son las fuentes para fortalecer los sumideros y embudos de conversión (puntos con mayor acceso) ya sean páginas oficiales, búsqueda desde YouTube u otros medios (mail, mensaje al móvil, redes sociales...)

Área geográfica	Visualizaciones	Minutos de visualización estimados	Duración media de las visualizaciones
España	5.107 (58%)	9.683 (55%)	1:53
México	1.093 (12%)	2.180 (12%)	1:59
Ecuador	453 (5,1%)	817 (4,7%)	1:48
Perú	374 (4,2%)	777 (4,4%)	2:04
Colombia	257 (2,9%)	652 (3,7%)	2:32
Estados Unidos	183 (2,1%)	286 (1,6%)	1:33
Chile	135 (1,5%)	277 (1,6%)	2:03
Venezuela	107 (1,2%)	245 (1,4%)	2:17
Brasil	96 (1,1%)	208 (1,2%)	2:10
Alemania	91 (1,0%)	288 (1,6%)	3:09

1/1/2015 – 31/1/2015



Tabla 23: Comportamiento del canal, segmentación por género y geografía durante el mes de Enero de 2015.

La importancia de la tabla adjunta se centra especialmente en la segmentación por edades. Podemos ver que el mayor volumen de usuarios es de 18 a 24 años, un dato coherente dado que coincide con la edad universitaria pero es de vital importancia potenciar la franja entre 13 y 17 años dado que los mayores consumidores de *Youtube* se encuentran en esa franja y son futuros estudiantes indecisos. *Youtube* es un canal excelente para mostrar la vida universitaria a los futuros estudiantes y conseguir un alto *engagement* con contenidos dinámicos mostrando la vida en un campus. A través de los *MOOCs* la *Universidad Autónoma de Madrid* busca acercar la universidad a los futuros estudiantes y sobre todo ofrecer un canal de divulgación internacional y de libre acceso. *Google* en su ranking de posicionamiento siempre da preferencia a los contenidos audiovisuales y la plataforma incluso permite la reproducción en streaming de eventos de importancia como la visita de Grimson a la UAM.



## 6.4 big data y redes sociales: Facebook



Twitter o Facebook se han convertido en el nuevo canal de comunicación para las universidades. Son el medio idóneo para dar a conocer novedades a los alumnos, fechas señaladas, resolver dudas inmediatas e incluso, impartir píldoras de conocimiento dinámicas. De hecho, Facebook nació como una red para promover la interacción entre alumnos universitarios en Estados Unidos.

El tipo de contenido debe fomentar el acercamiento hacia sus alumnos o futuros alumnos. La idea es que los alumnos puedan sentirse parte de una comunidad que los ligue a su centro de estudios tanto dentro como fuera del campus.

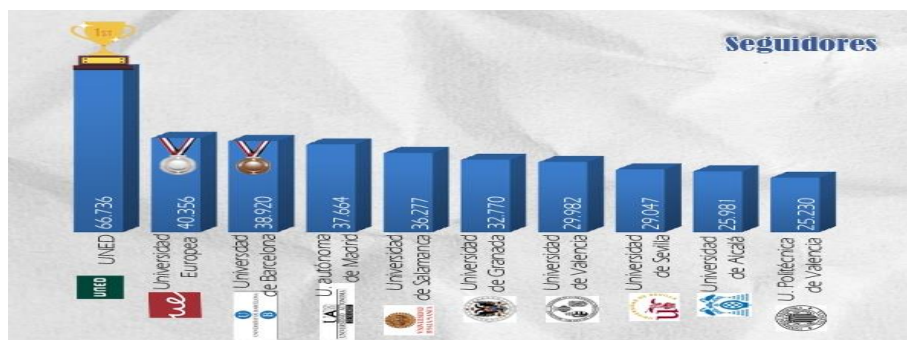


Figura 41: Top 10 de las Universidades españolas en Facebook en 2014.



#### Personas

**807** Personas que están hablando de esto

**46 638** Total de Me gusta de la página  
▲ 0,5% desde la última semana

Figura 42: Cabecera de la página institucional de Facebook de la UAM.

El grado de conversación e interacción que genera una página y por tanto su éxito puede medirse por el “número de personas que están hablando de esto” dividido por el “número de personas que les gusta esto”, cifras públicas que se encuentran en todas las páginas de Facebook.

Se recibe una media de 18 peticiones al día de información para la realización de estudios de postgrado en UAM.

Se atiende una media de 20 mensajes al día en relación a información administrativa o de posibles incidencias.

Existen infinidad de herramientas que permiten el procesado y análisis de los datos recogidos por Facebook. Se expone el análisis de alguna de las herramientas más populares con ejemplos aplicados sobre la página de la Universidad Autónoma de Madrid.

	<b>Dato</b>	<b>Ratio Diciembre 2014</b>
<b>Nuevos fans</b>	+1056 (Total 47279)	+34%
<b>Publicaciones</b>	34	-17%
<b>Me gusta</b>	3783	-36,6%
<b>Comentarios</b>	287	-47%
<b>Compartidos</b>	540	-3%
<b>Mensajes privados</b>	70 (Total: 1153)	-19%

Tabla 24: Resumen general de la página de Facebook por usuarios e interacción.



## Análisis en Facebook con LikeAlyzer.



Muestras las métricas de fans, crecimiento de fans y engagement y los post por día y de qué tipo son (texto, link, fotos, videos, preguntas...). LikeAlyzer te dice el timing, es decir, la hora a la que se realizan más actualizaciones en la página.

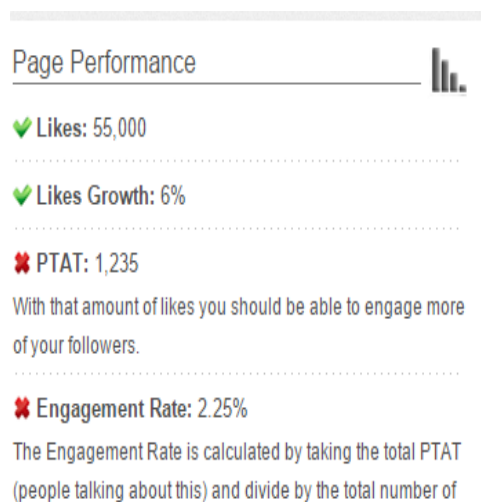


Tabla 25: Page Performance y Post by Pages obtenidos con LikeAlyzer.

LikeAlyzer calcula el número de post diarios de la página, el engagement, el horario en el que se producen más interacciones y el tiempo de respuesta.

Para evaluar la posición relativa de tu marca frente a la de tu competencia, detectar qué tácticas está desarrollando y saber cuáles le están resultando más eficaces.

En relación a los ratios anteriores se analiza la posición de la Universidad Autónoma de Madrid. La posición 45 ocupada por la UAM a pesar de tener más Likes (seguidores) que la Universidad Pompeu Fabra se debe al alto índice de engagement. Google y sus algoritmos siempre dan preferencia a las redes sociales con mayor interacción, es decir es más importante tener una comunidad menor pero activa

que una comunidad de miles de seguidores que no compartan el contenido ni interactúen con la marca (seguidores fantasmas).

#	Page	Likes	PTAT	ER	LikeRank
41	 <b>Universitat Pompeu Fabra</b> University	25,733	1,276	4.96%	70
42	 <b>Universidad Europea</b> University	66,778	574	0.86%	70
43	 <b>Escuela de Educación y Turismo de Ávila</b> University	142	53	37.32%	69
44	 <b>CETT - UB Formació i Recerca en Hoteleria i Turisme</b> University	11,206	157	1.4%	69
45	 <b>Universidad Autónoma de Madrid</b> University	55,000	1,235	2.25%	69

Figura 43: Ranking page de Universidades españolas en Facebook por LikeRank.

En relación a las páginas bajo análisis de la Universidad Autónoma de Madrid vemos que el crecimiento es positivo y presentamos el mayor crecimiento en número de seguidores en Enero 2015.

Página	Total de Me gusta de la p	Nuevos Me gusta	Publicaciones de es	Interacción de esta semana
1  <b>Harvard University</b>	4,1m 	▲0,3% desde la semana pasada	19	41,4K 
2  <b>UNED - Universidad Nac...</b>	84,5K	▲0,4% desde la semana pasada	6	971
3  <b>Universidad Europea</b>	51,9K	▲0,2% desde la semana pasada	15	273
4  <b>Universidad de Granada</b>	48,3K	▲0,5% desde la semana pasada	61	1,4K
5  <b>Universidad Autónoma ...</b>	48,2K	▲0,6% desde la semana pasada	18	1K
6  <b>Universidad Compluten...</b>	40,1K	▲0,3% desde la semana pasada	0	0
7  <b>Universidad Carlos III d...</b>	26,6K	▲0,4% desde la semana pasada	12	391

Figura 44: Ranking page de universidades bajo estudio en 2015 internacionales y privadas.

Los datos que se pueden obtener de una publicación universitaria se muestran en la imagen adjunta. Podemos ver personas alcanzadas (personas que ven la publicación) y el desglose de la actividad realizada con dicha publicación.

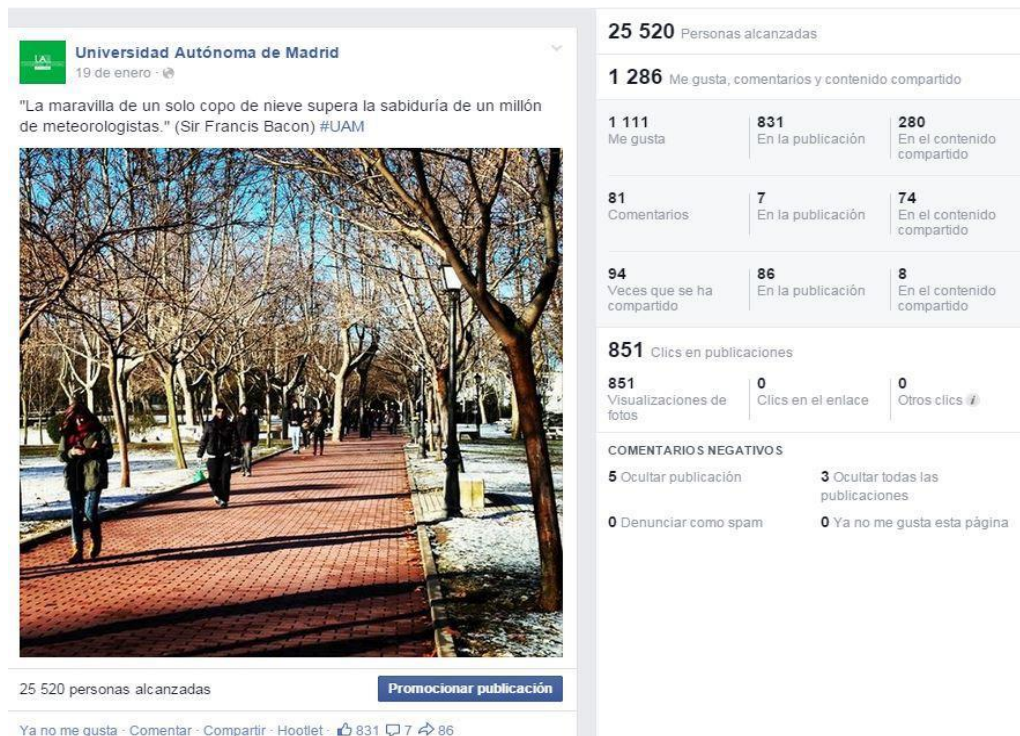


Figura 45: Análisis de los datos obtenidos de una publicación en Facebook.

En la imagen posterior podemos observar el alcance de la publicación en términos de posicionamiento del buscador *Google*. Podemos ver el número de personas alcanzadas y la fuente de ese tráfico. Tráfico "orgánico" (*SEO*, procedente del buscador), tráfico de pago (*SEM*, procedente de *Google Adwords*).

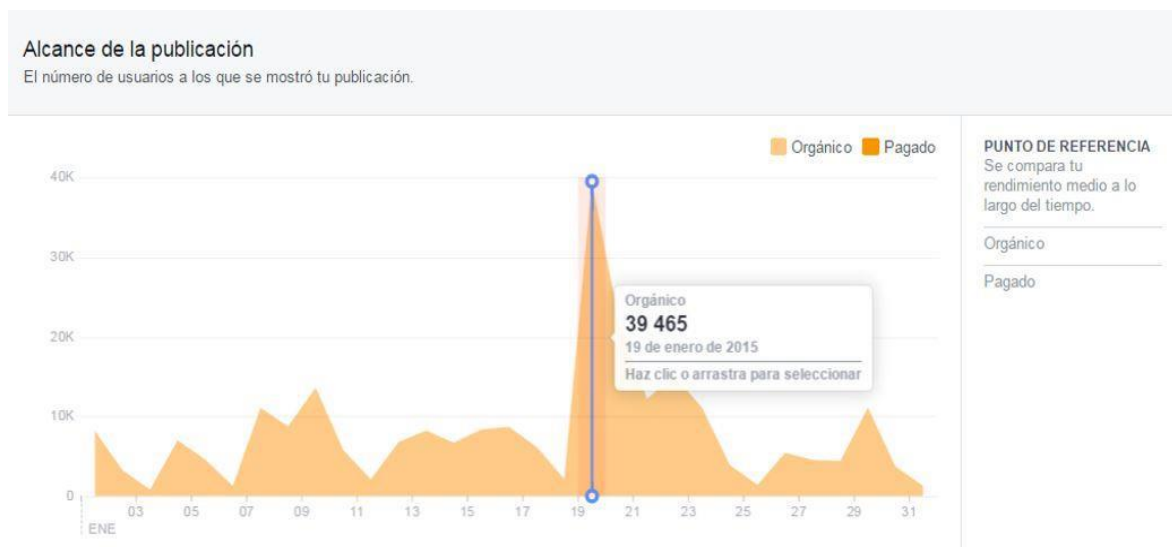




Figura 46: Análisis de interacción de la comunidad en Facebook.

Por último segmentamos por completo la comunidad por géneros y edades al igual que realizamos en Twitter. El género es un elemento determinante dado que existen redes sociales que tienen mayor interés para determinados sexos. En este caso vemos que existe una supremacía del sector femenino y vemos una fuerte deficiencia de seguidores en el sector comprendido entre los 13 y 17 años absolutamente diferente con el dato arrojado por *Youtube*. Se adjunta también el análisis geográfico y por idioma de los usuarios.

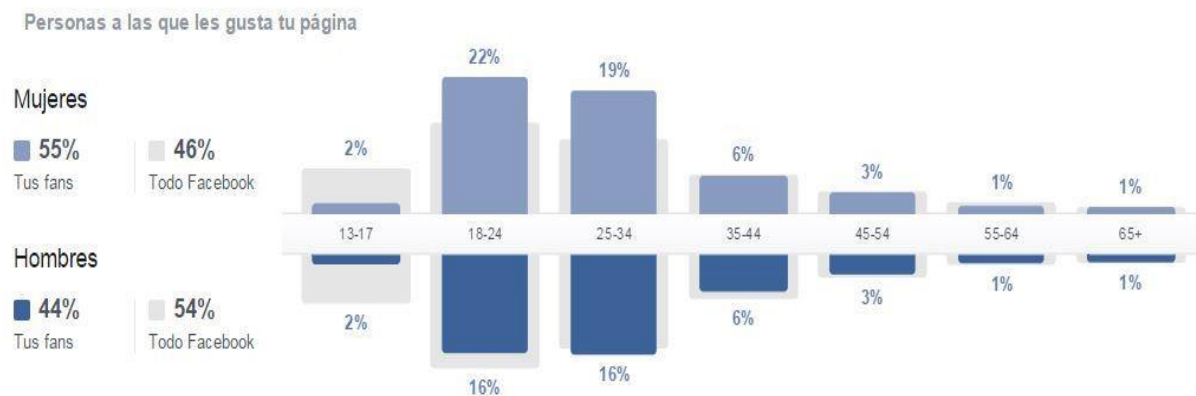


Figura 47: Infografía por sexo y edad de la comunidad de usuarios UAM en Facebook.

Pais	Tus fans	Ciudad	Tus fans	Idioma	Tus fans
España	20 054	Madrid	13 372	Español	20 677
México	6 144	Lima, Perú	2 023	Español (España)	15 255
Perú	3 659	México, D. F., México	1 704	Inglés (EE.UU.)	3 746
Colombia	1 587	Barcelona, Cataluña	522	Inglés (Reino Unido)	1 690
Brasil	1 367	Bogotá, Colombia	463	Francés (Francia)	1 617
Argentina	1 123	Guadalajara (México), Jali...	376	Portugués (Brasil)	1 274
Estados Unidos de América	1 095	Buenos Aires, Argentina	356	Italiano	978
Italia	1 006	Santiago de Chile, Chile	315	Portugués (Portugal)	440
Venezuela	859	Monterrey, Nuevo León, M...	289	Árabe	436
Chile	631	Puebla de Zaragoza, México	284	Alemán	369
Ecuador	612	Londres, Inglaterra, Reino ...	276	Polaco	301
Reino Unido	581	Alcobendas, Madrid	252	Indonesio	172
Costa Rica	538	Caracas, Distrito Capital, ...	226	Turco	166

**Figura 48: Análisis geográfico de los usuarios de Facebook.**

## 6.5 Big Data: Analítica Web: Google Analytics y cookies.

*“Recopilación, medición, evaluación y explicación racional de los datos obtenidos de Internet, con el propósito de entender y optimizar el uso de la página web de la organización”.*

### **Asociación Española de Analítica Web**

La analítica web surgió al mismo tiempo que las propias páginas web al inicio de la década de los noventa con la compañía *Webtrends*. Hoy en 2015 el 90% de los sitios web en España utilizan *Google Analytics*.

En Internet se puede medir todo y, casi a tiempo real. Los datos son muy variados pero se pueden dividir inicialmente en:

- **Logs vs Tags**

1. En la medición por Logs, al navegar se envía información al servidor que se almacena en archivos .log.

Una IP es un visitante único: permite medir cuando se descarga un archivo, tiene como fundamento que una página descargada no se mide, poca flexibilidad para manejar grandes volúmenes de datos ya que puede demorarse días los resultados de su análisis.

2. Los Tags, son trozos de código Javascript colocados dentro una de una página web que envían información a un servidor que la procesa.

Un ordenador es un visitante único: permite medir el clic si la descarga fue completa, cada página descargada sí se mide, flexible y a tiempo casi real.

- **Censales vs Panel**

1. Los paneles estudian a un grupo reducido de usuarios en detalle y a partir de ellos hacen estimaciones para el resto de usuarios
2. Los sistemas censales analizan a todos los usuarios directamente. Ofrecen datos. Dado el elevado volumen de usuarios en Internet, requiere de una muestra con 20.000 usuarios para que resulte representativa.

## **La ventaja de los medios digitales: Cookies.**

Las cookies son un archivo descargado en el equipo de un usuario con la finalidad de almacenar y recuperar datos. Podemos agruparlas según su finalidad.

*Cookies técnicas:* permiten al usuario la navegación y la utilización de las diferentes opciones o servicios que en ella existan como, por ejemplo, identificar la sesión.

*Cookies de personalización:* Permiten al usuario acceder al servicio con algunas características predefinidas en función de una serie de criterios en el terminal del usuario como por ejemplo el idioma.

*Cookies de análisis:* Permiten el seguimiento y análisis del comportamiento de los usuarios de los sitios web a los que están vinculadas.

*Cookies publicitarias:* Permiten la gestión, de la forma más eficaz posible, de los espacios publicitarios en base a criterios como el contenido editado o la frecuencia en la que se muestran los anuncios.

*Cookies de publicidad comportamental:* Permite desarrollar un perfil específico de usuario para mostrar publicidad en función del mismo.

## **Métricas básicas (UUs, PVs, Visitas, % rebote, tiempo promedio)**

- *Visitantes únicos:* Número de visitantes no duplicados que han accedido al sitio web. Se genera una cookie que es única y que se utilizará para identificarlo.
- *Visitas:* Una visita es un conjunto de interacciones que tienen lugar en un sitio web durante un periodo determinado (30 minutos en *Google Analytics*) momento en el que se elimina la cookie `__utmb`. Pero esta se ve modificada con cada evento y, por tanto, cada una de las solicitudes adicionales restablecerá la caducidad de la cookie en 30 minutos. La otra forma de finalización forzosa de la cookie `__utmb` se produce por la finalización del día.



- *Páginas Vistas:* Son las páginas a las que se accede dentro del sitio web. Las páginas compuestas por cantidad de respuestas de un servidor a la petición de un navegador.
- *Duración media de la visita:* Es la duración media de una sesión.
- *Porcentaje de rebote:* Porcentaje de visitas de una sola página y en las que el usuario ha abandonado el sitio en su página de entrada.
- *Páginas de destino* (aquellas de nuestro sitio a las que llegan los usuarios) páginas de salida (la última página visitada antes de que el usuario abandone nuestro sitio).



## 6.5.1 Analítica Web: SEO y SEM.

SEO que es el Search Engine Optimization, haciendo referencia al posicionamiento en buscadores de forma orgánica (sin pago). El SEO busca posicionamiento y tráfico para nuestro sitio web.

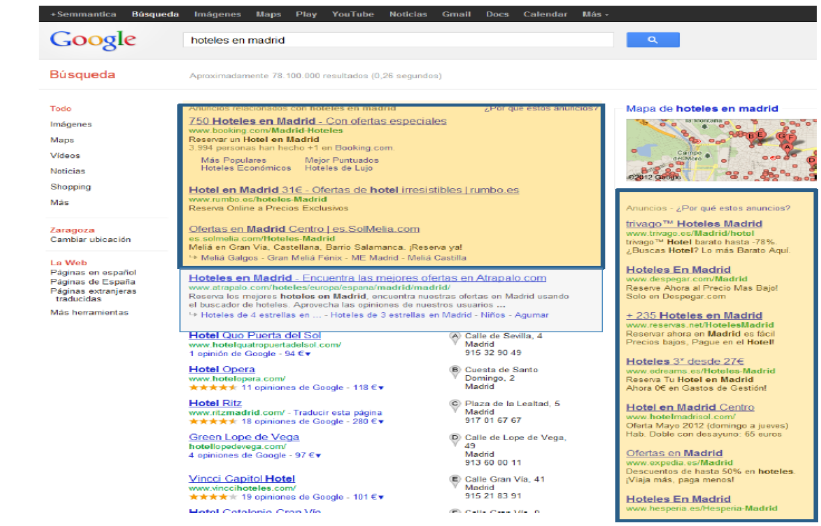


Figura 49: Imagen del posicionamiento SEO de una página web.

SEM (Search Engine Marketing) orientado a conseguir estar bien posicionados por encima de los rivales para momentos determinados. Está relacionado con Google AdWords, la herramienta de Google para invertir y comprar palabras para conseguir el posicionamiento destacado

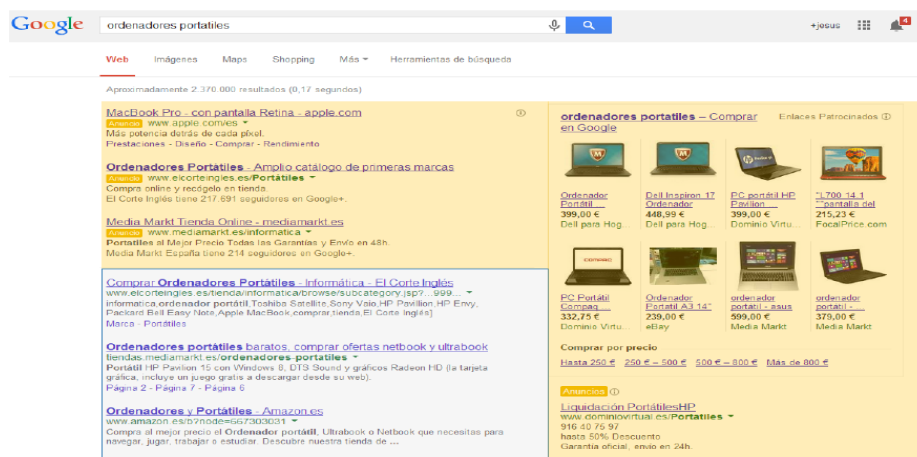


Figura 50: Imagen del posicionamiento SEM de una página web.

- Tráfico directo (el usuario llega directamente tecleando la URL en el navegador).
- Tráfico de referencia (el usuario llega siguiendo un vínculo desde otro sitio). Las redes sociales están incluidas aquí aunque luego tengan un sitio específico donde analizarlas con mayor detalle.

Las *keywords* son la palabra o combinación de palabras con las que los algoritmos de los buscadores clasifican y categorizan nuestra página web en su base de datos.

*Google Trends Explore* es una herramienta de Google que muestra los términos de búsqueda más populares del pasado reciente. *En Google Trends Explore* podemos ver con cuánta frecuencia se realiza una búsqueda particular en cualquier región del mundo y en varios idiomas.



Figura 51: Report de Google Trends para Universidad Autónoma de Madrid.

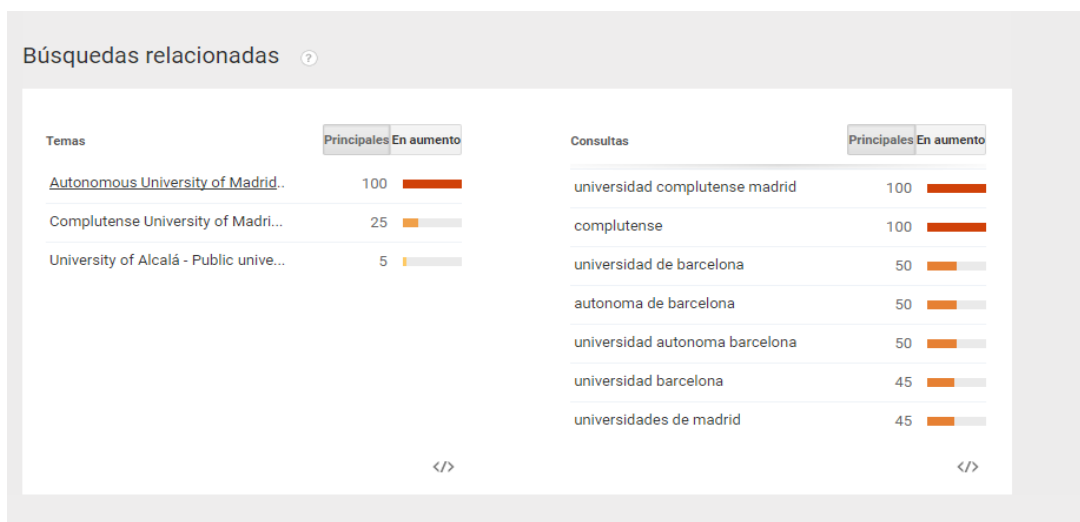


Figura 52: Report de Google Trends con consultas frecuentes para Universidad Autónoma de Madrid.

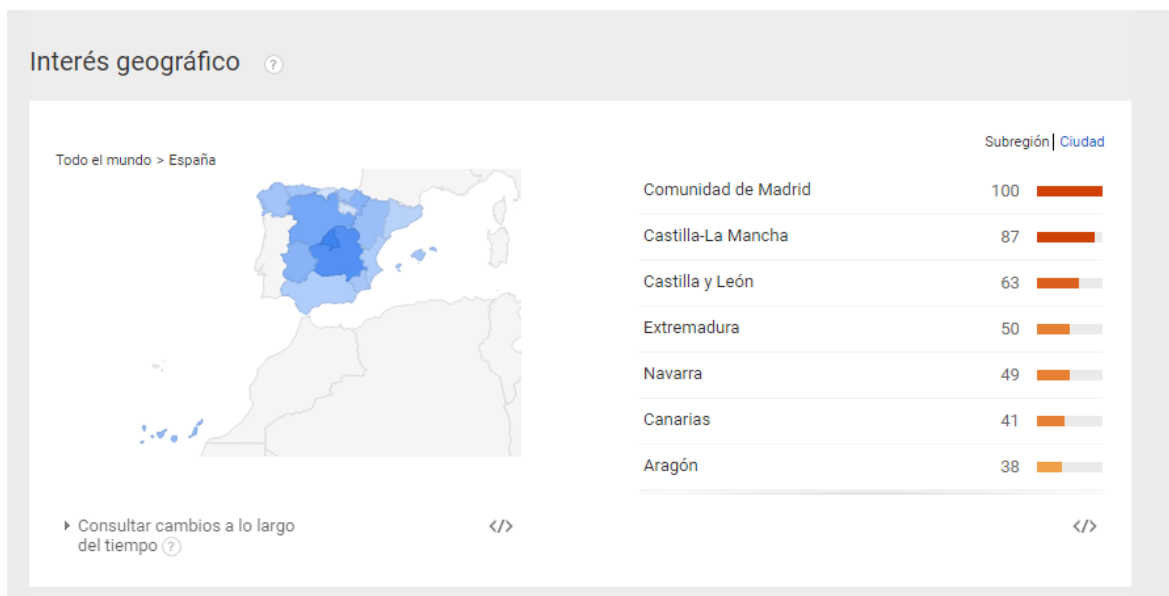


Figura 53: Report de Google Trends con geolocalización del interés por “Universidad Autónoma de Madrid”.

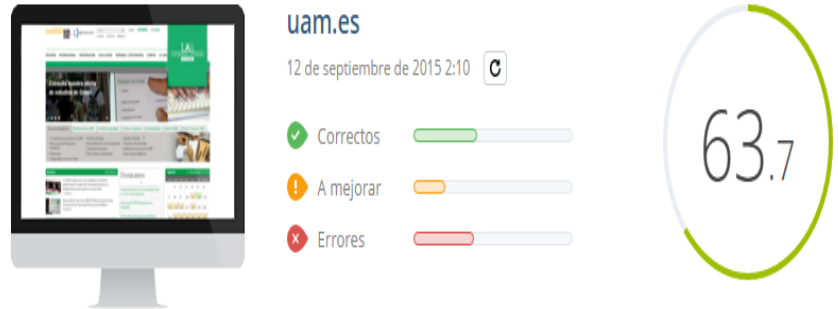
Podemos identificar las palabras clave con mayor rendimiento, ver con qué palabras clave están accediendo los usuarios a nuestro sitio y así generar más contenido o modificarlo para adecuarlo a estas búsquedas.

### Proceso de optimización para buscadores SEO:

- 1) Análisis Inicial Web
- 2) Análisis Palabras Claves
- 3) Análisis de la Competencia
- 4) Análisis de Enlaces Entrantes
- 5) Alta en Buscadores y Directorios
- 6) Trabajo en Redes Sociales
- 7) Mejoras de la Web (Contenido, estructura, responsive)
- 8) Reportes de Evolución

## 6.5.2 Analítica Web: Woorank valoración onpage

WooRank es la mejor herramienta online para realizar un análisis SEO (tráfico), social media (datos sobre el impacto en las redes sociales desde la URL) y optimización dispositivos móviles, tiempo de carga de web y visualización en dispositivos móviles.



Analizamos la analítica SEO de la URL [www.uam.es](http://www.uam.es) con aspectos como los títulos, descripciones, imágenes, marcos y páginas interna. Woorank describe según sus parámetros un pagerank 63.7 siendo el pagerank una valoración de Google de 0 a 100 en función de parámetros como popularidad, tiempo de carga o peso de las imágenes.

### SEO

	Etiqueta Título	Universidad Autónoma de Madrid
		Longitud: 30 caracteres
	Meta Descripción	Universidad Autónoma Madrid
		Longitud: 27 caracteres
	Vista previa de Google	<a href="http://www.uam.es/">Universidad Autónoma de Madrid</a> uam.es/ Universidad Autónoma Madrid

✓ Encabezados  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

<H1>	<H2>	<H3>	<H4>	<H5>
1	15	0	0	0

<H1>
<H2> Eres de la AutónomaÚnete
<H2> I Edición del Concurso deEnsayo Científico de la UAMPresentación de trabajos del 26 al 30 de octubre de 2015
<H2> Campus sosteniblesituado en plena naturaleza
<H2> Conoce nuestros programas de movilidad internacional
<H2> Nuevos estudiantes
<H2> Publicaciones UAM
<H2> Unidad de Igualdad
<H2> Cultura y Deporte
<H2> Empleabilidad
<H2> Alumni UAM
<H2> Redes Sociales UAM
<H2> Noticias
<H2> Cultura Científica
<H2> Destacamos
<H2> Agenda

En este apartado analizamos la facilidad de la página para ser indexada y encontrada por los usuarios y motores de búsqueda. El tamaño del título se considera aceptable (27 caracteres autodefinidos) pero la definición en metadatos se basa en repetir el título por lo que se pierde la capacidad de ampliar audiencia.

Las cabeceras están basadas en las últimas publicaciones de la UAM siendo una técnica poco beneficiosa a largo plazo para futuros estudiantes que no conocen el campus y realizan búsquedas de palabras como “grado”, “medicina” más generalistas.

Nube de palabras clave

uam 15   sociales 6   prácticas 5   redes 4   cultura 4   igualdad 4  
 universidad 4   externas 3   campus 3   claustro 3

Coherencia de palabras clave

	Palabras clave	Contenido	Título	Descripción	<H>
★★★	uam	15	×	×	✓
★★★	sociales	6	×	×	✓
★★★	prácticas	5	×	×	×
★★★	redes	4	×	×	✓
★★★	cultura	4	×	×	✓
Palabras clave (2 términos)		Contenido	Título	Descripción	<H>
★★★	redes sociales	4	×	×	✓
★★★	prácticas externas	3	×	×	×
★★★	prácticas remuneradas	2	×	×	×
★★★	sociales centros	2	×	×	×
★★★	uam unidad	2	×	×	×
Palabras clave (3 términos)		Contenido	Título	Descripción	<H>
★★★	prácticas externas trámites	2	×	×	×
★★★	universidad autónoma madrid	2	✓	✓	×
★★★	uam prácticas externas	2	×	×	×

En relación a lo anterior vemos que en las palabras clave ( palabras más usadas en nuestro dominio) casi ninguna aparece en las cabeceras (H) y ni si quiera en la descripción. Es especialmente importante el fallo en la palabra “uam”.

Alt Attribute

Se han encontrado 34 imágenes en esta URL.  
 6 atributos ALT están vacíos o no aparecen.

- [http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/punto\\_verde.png](http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/punto_verde.png)
- [http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/punto\\_blanco.png](http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/punto_blanco.png)
- [http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/icono\\_pause.png](http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/icono_pause.png)
- [http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/icono\\_play.png](http://uam.es/StaticFiles/UniversidadAutonomaMadrid/img/icono_play.png)

Ratio texto/HTML

8,2%

Páginas indexadas

166.457


Google+ Publisher

Su sitio no tiene etiqueta rel="Publisher" que le permita ser asociado a una página Google+.

Presentamos un bajo ratio de HTML por ello tenemos mala adaptación con dispositivos móviles al no utilizar HTML5. Se debe al uso de la aplicación Fatwire sin implementar el paquete CS.

**Enlaces en página** Se han encontrado un total de 114 enlace(s), incluyendo 1 enlace(s) a archivos

ⓘ ⓘ ⓘ ⚙ ⚙ ⚙



- Enlaces externos: Nofollow ( 0% )
- Enlaces externos: Follow ( 49.8% )
- Enlaces internos ( 50.2% )

Anchor text	Tipo	Follow
Programa de prácticas remuneradas de posgrado	Enlaces externos	Follow
Santander Universidades	Enlaces externos	Follow
Acceso a Sede	Enlaces externos	Follow
Universia	Enlaces externos	Follow
Directorio	Enlaces externos	Follow
Biblioteca	Enlaces externos	Follow

Mostrar más

---

**Enlaces rotos** Hemos detectado 2 enlaces rotos en su página web.

ⓘ ⓘ ⓘ ⚙ ⚙ ⚙

URL	Estado	Code
http://uam.es/ss/Satellite?c=UAM_Noticia_FA&language=es&page...	Not Found	404
http://uam.es/ss/Satellite?c=UAM_Evento_FA&language=es&page...	Not Found	404

---

**Resolver www** ¡Cuidado! No tiene configurada una redirección 301 para enviar tráfico a su dominio preferido. Si las versiones con www. y sin www. cargan correctamente ¡es contenido duplicado!

ⓘ ⓘ ⓘ ⚙ ⚙ ⚙

---

**Canonicalización de la IP** Sí

ⓘ ⓘ ⓘ ⚙ ⚙ ⚙

---

**Robots.txt** No disponible

ⓘ ⓘ ⓘ ⚙ ⚙ ⚙

En este apartado vemos que la mayoría son enlaces externos, la página de la UAM no genera como tal contenido sino que lo enlaza a otras páginas sirviendo en definitiva como un portal de hipervinculos más que un portal de contenido propio (50% de su contenido es externo).

✖ Mapa del sitio XML *No disponible*  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

✖ Reescribir URL ¡Cuidado! Se han detectado parámetros en varias URL.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

✖ Guiones bajos en su URL Hemos encontrado guiones bajos en esta URL y/o en las URLs de sus páginas internas.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

`http://uam.es/ss/Satellite/es/1233310431942/subHome/Personal_y_estudiantes.htm`  
`http://uam.es/ss/Satellite/es/1233310431945/subHome/La_UAM.htm`  
`http://uam.es/ss/Satellite/es/123331043220... /10_razones_para_estudiar_en_la_UAM.htm`  
`http://uam.es/ss/Satellite/es/1233310432206/sinContenido/Estudios_de_Grado.htm`  
`http://uam.es/ss/Satellite/es/1242651583... /Oficina_para_la_Participacion_Estudiantil.htm`  
`http://uam.es/ss/Satellite/es/1242652010877... /Area_de_Atencion_a_la_Discapacidad.htm`

Mostrar más

✖ Factores limitantes **✖ Flash: Sí**  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙ **✓ Marcos: No**

✓ Blog Se ha encontrado un blog en este sitio web.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

## Móvil

✖ Optimización Móvil **Poor**  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙  
Esta web no está muy optimizada para dispositivos móviles  
✖ CSS para móvil ✖ Redirección Móvil



✖ Pantalla Táctil Asegúrese de que los elementos táctiles más importantes sean lo bastante grandes como para tocarlos sin dificultad.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

✓ Compatibilidad Móvil Perfecto, ningún objeto incorporado ha sido detectado.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

✖ Tamaño de Fuente El texto de esta página web es demasiado pequeño para ser legible en un dispositivo móvil.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙

✖ Ventana Gráfica ✖ Esta página no especifica ninguna ventana gráfica, o ésta ventana gráfica no está bien configurada.  
⊙ ⊙ ⊙ ⊙ ⊙ ⊙  
✓ El contenido encaja en la ventana gráfica especificada.



Velocidad del Dispositivo Lento

- ✗ Eliminar el JavaScript que bloquea la visualización y el CSS del contenido de la mitad superior de la página
- ✗ Habilitar compresión
- ✗ Especificar caché de navegador
- ✗ Minificar JavaScript
- ✗ Optimizar imágenes

La versión móvil de su web debe mostrar el contenido de la mitad superior de la página en menos de 1 segundo. Esto permite a los usuarios interactuar con la página de manera inmediata. Puesto que la CPU de un dispositivo móvil es menos poderosa que la CPU de un ordenador, es necesario mejorar la velocidad para reducir el consumo de la CPU (por ejemplo JavaScript Parse time).

Echen un vistazo a las Reglas de Google PageSpeed Insights para saber cómo optimizar cada criterio de esta sección.

Entornos Móviles No detectamos ningún entorno móvil.

Google penaliza las páginas con mayor retardo en carga para mejorar la experiencia de usuario, una de las razones de ese retraso son las imágenes cargadas como jpeg que introducen cierta latencia por ello se ofrece una serie de soluciones posibles entre ellas cambiar el hosting.

Desde 21 de Abril de 2015, los sitios web que no estén optimizados para dispositivos móviles han perdido posicionamiento en los resultados de búsqueda de Google realizadas desde estos dispositivos.

## Usabilidad

● URL	http://uam.es Longitud: 3 caracteres														
● Favicon	Perfecto, la página web tiene favicon.														
✓ Página 404 Personalizada	¡Perfecto! Su sitio web tiene una página personalizada de error 404.														
● Tamaño de página	41,3 Kb (el promedio mundial es 320 Kb)														
✓ Tiempo de carga	0,41 segundo(s) (101,28 kB/s)														
✓ Idioma	Declarado: <i>Español</i> Detectado: <i>Español</i>														
✗ Marcado de Datos Estructurados	Ningún Marcado de Datos Estructurados ha sido detectado														
● Disponibilidad del dominio	<table><thead><tr><th>Dominios (TLD)</th><th>Estado</th></tr></thead><tbody><tr><td>uam.com</td><td>Caduca en 2 años</td></tr><tr><td>uam.net</td><td>Caduca en 2 años</td></tr><tr><td>uam.org</td><td>Este dominio está registrado</td></tr><tr><td>uam.info</td><td>Este dominio está registrado</td></tr><tr><td>uam.biz</td><td>Caduca en 7 meses</td></tr><tr><td>uam.eu</td><td>Este dominio está registrado</td></tr></tbody></table>	Dominios (TLD)	Estado	uam.com	Caduca en 2 años	uam.net	Caduca en 2 años	uam.org	Este dominio está registrado	uam.info	Este dominio está registrado	uam.biz	Caduca en 7 meses	uam.eu	Este dominio está registrado
Dominios (TLD)	Estado														
uam.com	Caduca en 2 años														
uam.net	Caduca en 2 años														
uam.org	Este dominio está registrado														
uam.info	Este dominio está registrado														
uam.biz	Caduca en 7 meses														
uam.eu	Este dominio está registrado														

## Visitantes

✓ Estimación del tráfico	Muy Alto
--------------------------	----------

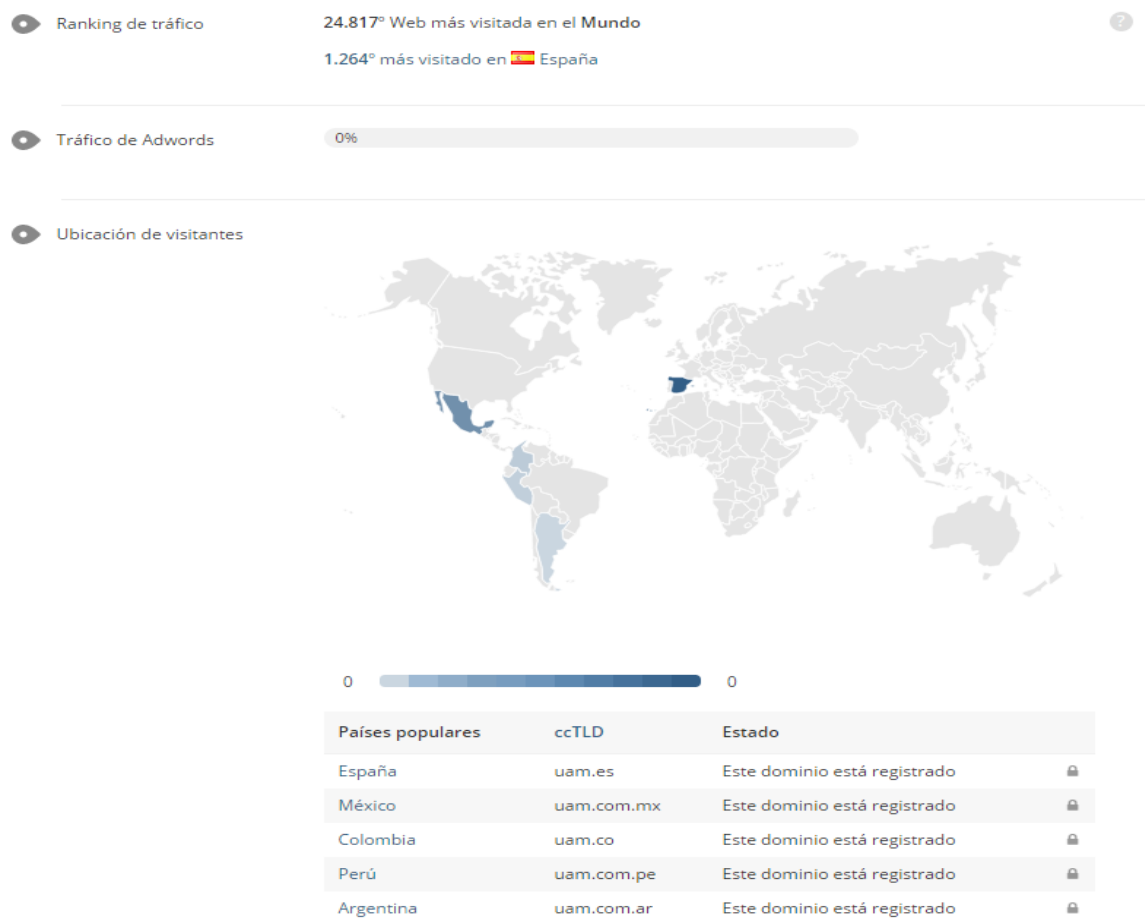


Figura 54: Mapa geográfico del tráfico web de la página de la Universidad Autónoma de Madrid.

### 6.5.3 SEO para apps: ASO (App Store Optimization).

**ASO (App Store Optimization)** es el proceso de optimización de las aplicaciones móviles para mejorar el posicionamiento de las aplicaciones móvil en los ranking de marketplace, para conseguir un mayor número de descargas.

Los usuarios consumen un 80% del tiempo en aplicaciones web. Nacen 2700 aplicaciones nuevas todos los días

*Estrategia ASO en Google Play.*

Google Play por tratarse de un producto de Google tiene acceso a todos los índices de búsqueda recogidos por el principal buscador

- **Título de la App:** La mejor estrategia para el posicionamiento ASO es un título corto que explique para qué sirve tu aplicación.

COMBINATIONAL CIRCUITS - DSLAB



Figura 55: App Comb Circuits DSLab

- **Keywords:** las palabras que harán que aparezcas en los resultados. Google “penaliza” el uso “repetitivo o irrelevante” de keywords en el título y en la descripción.
- **Descripción de la App:** explicar qué es tu aplicación, qué hace, y cuáles son sus beneficios. El límite de caracteres en estas descripciones es de 4000.

Combinational Circuits, es una aplicación educativa que ofrece una serie de ejercicios de electrónica combinacional.

El principal objetivo es que el usuario pueda estudiar y resolver problemas desde su dispositivo android.

La aplicación indica si se ha resuelto el ejercicio correctamente y además incluye consejos que ayudan a la resolución de los problemas.

Se pueden guardar soluciones de los ejercicios, enviar una solución a otra persona por correo y que esta la cargue en su aplicación "Combinational Circuits".

Se incluye una ayuda (help) con el manejo básico de la aplicación. También está disponible un tutorial sobre circuitos combinacionales.

Figura 56: Descripción App Comb Circuits DSlab

- **Tipo de App:** En Google Play sólo hay dos opciones: Aplicaciones y Juegos.  
**Categoría:** Define correctamente tu producto para una categoría específica.



Figura 57: Categoría y tipo de app Comb Circuits DSlab

- **Capturas de pantalla:** La mayoría de las decisiones se toman gracias a las sensaciones visuales.

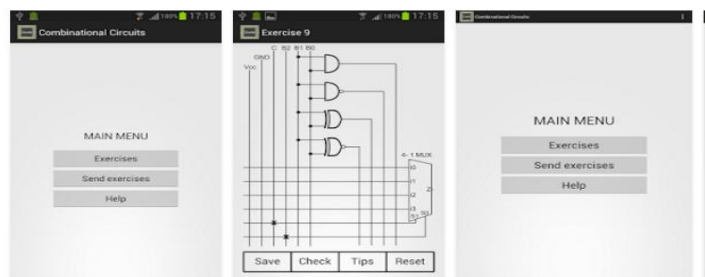


Figura 58: Capturas de la app Comb Circuits DSlab

- **Demostración de App en YouTube:** subir un video en YouTube sobre el funcionamiento de tu aplicación de un minuto.
- **Valoraciones y votos en la aplicación:** La posición en la primera página de Google Play es una combinación de opiniones, valoraciones y cuanta más gente instale una aplicación más visibilidad en Google Play.

OPINIONES

 Escribir una opinión

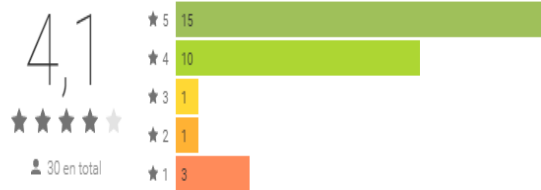


Figura 59: Valoración de la app Comb Circuits DSLab

- **Linkbuilding:** cuantos más enlaces desde sitios web externos más alto nos posiciona Google.

<https://dslabuam.wordpress.com/>

<https://www.facebook.com/Digital-System-Lab-1379095145745111/>

## 6.5.4 Uso de blogs universitarios: Edublogs

Los Edublogs son aquellos blogs que se usan con fines educativos en entornos de aprendizaje, permitiendo compartir información y opinión acerca de las asignaturas. Para el alumnado es también un foro de opinión y un espacio para poder medir sus progresos mediante aplicaciones o técnicas de autoevaluación.

Ejemplos de blog de docentes:

Tíscar: <http://tiscar.com/>

Educación y TIC: <http://domingomendez.blogspot.com/>

**WORDPRESS** Es el sistema que más desarrollo está teniendo en los últimos años y la mejor opción que existe actualmente en PHP. Altamente personalizable con capacidad de diseñar gran variedad plugins y widgets.

<https://dslabuam.wordpress.com/>



**DIGITAL SYSTEM LABORATORY**  
Escuela Politécnica Superior - Universidad Autónoma de Madrid

INICIO APPS ON ANDROID APPS ON IOS ACADEMIC PUBLICATIONS MEMBERS ACERCA DE

### COMBINATIONAL CIRCUITS - DSLAB

Comb

Get it on Google play

#### Descripción

Combinational Circuits, es una aplicación educativa que ofrece una serie de ejercicios de electrónica combinacional. El principal objetivo es que el usuario pueda estudiar y resolver problemas desde su dispositivo android. La aplicación indica si se ha resuelto el ejercicio correctamente y además incluye consejos que ayudan a la resolución de los problemas. [\(más...\)](#)

Posted in Apps on Android, Circuitos Electrónicos Digitales on 23 marzo, 2015. [Deja un comentario](#)

UAM UNIVERSIDAD AUTÓNOMA DE MADRID

Escuela Politécnica Superior

DSL<sup>ab</sup>

#### FOLLOW US

Introduce tu dirección de correo electrónico para seguir este Blog y recibir las notificaciones de las nuevas publicaciones en tu buzón de correo electrónico.

Figura 60: Landing page de Digital System Laboratory.

Al igual que en apartados anteriores se puede realizar un análisis completo de las métricas del blog creado para **DSL**Lab. Este espacio ha sido creado con el objetivo de ser un canal de comunicación del profesorado con los alumnos para hacer llegar al estudiante de una forma dinámica las últimas novedades en aplicaciones móviles desarrolladas por el **DSL**Lab. Se ha implementado a su vez la posibilidad de suscribirse a una *newsletter* del **DSL**Lab y un widget de chat para la comunicación directa de los usuarios para formular preguntas de forma anónima o mediante sus redes sociales.

Podemos conocer mediante esta herramienta cuales han sido los enlaces con mayor tasa de clicks, geolocalización de nuestros visitantes y la interacción con las publicaciones.

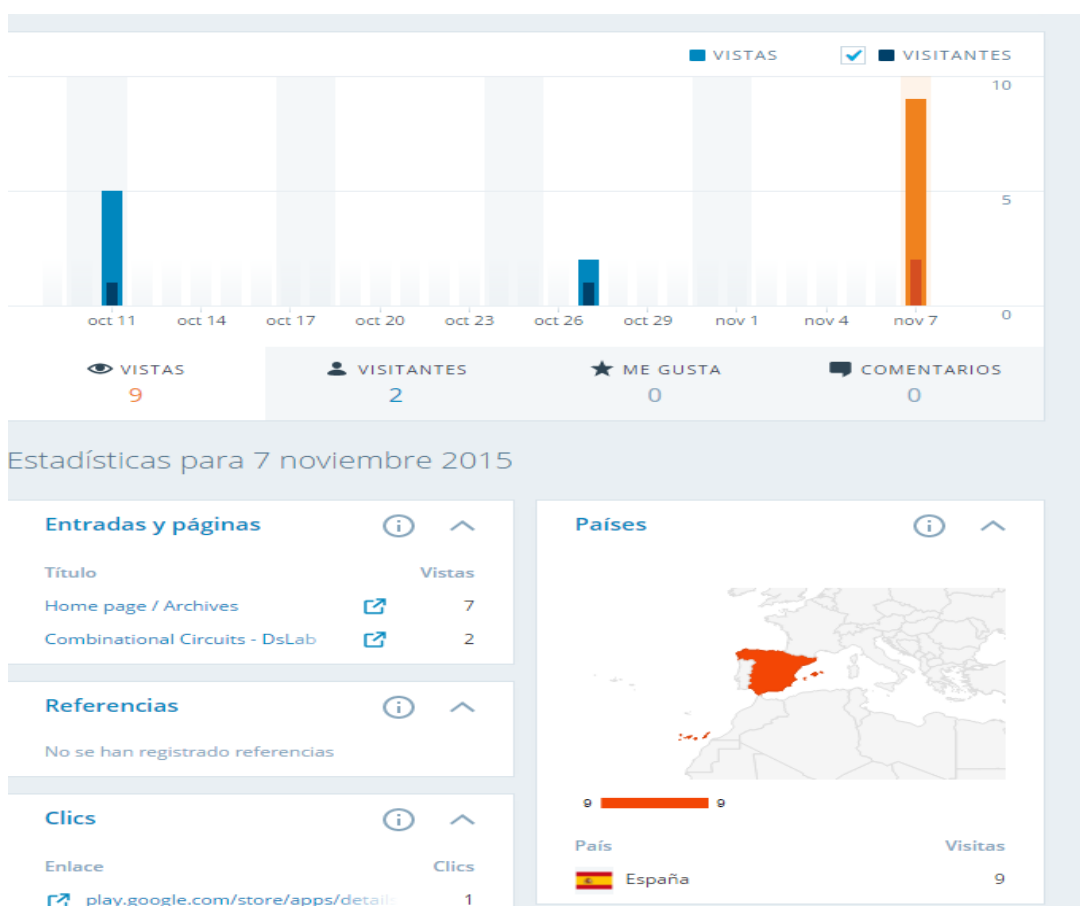


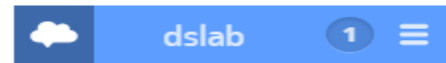
Figura 61: Estadísticas a 7 de noviembre de tráfico y posicionamiento.



**Seminarios y Cursos  
Android**  
16/03/15

<http://goo.gl/2E7qCU>

## CHAT DE CUESTIONES



ismixu14  
Buenos días tengo una consulta



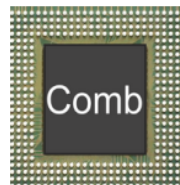
## EVENTOS PRÓXIMOS

No existen eventos próximos

También se ha creado un espacio para notificar los posibles nuevos eventos tales como salidas de proyectos o seminarios

Los blogs nos permiten por tanto aumentar el posicionamiento ASO de nuestras aplicaciones abriendo un nuevo canal de difusión. Se adjunta imagen de una de las pruebas:

## COMBINATIONAL CIRCUITS – DSLAB



### Descripción

Combinational Circuits, es una aplicación educativa que ofrece una serie de ejercicios de electrónica combinacional. El principal objetivo es que el usuario pueda estudiar y resolver problemas desde su dispositivo android. La aplicación indica si se ha resuelto el ejercicio correctamente y además incluye consejos que ayudan a la resolución de los problemas. Se pueden guardar soluciones de los ejercicios, enviar una solución a otra persona por correo y que esta la cargue en su aplicación "Combinational Circuits". Se incluye una ayuda (help) con el manejo básico de la aplicación. También está disponible un tutorial sobre circuitos combinacionales. Aplicación gratuita y offline. Contenidos: - Memoria ROM - Multiplexores - Puertas lógicas - PALS - BCD 7-segmentos - Codificador y decodificador - Código Gray - Karnaugh .



## Conclusiones y trabajo futuro

---

Big Data es una innovación pero también una nueva tecnología integradora que permite dar el impulso necesario en algunos ámbitos como el educativo o la relación con el cliente en los que la cantidad de datos y la falta de recursos han supuesto la toma de decisiones y procesos como dogmas. La información y las soluciones innumerables pueden suponer una nube ruidosa de que no permiten marcar con exactitud la solución que queremos de Big Data cayendo a veces en el error de considerar la solución más cara la única viable por falta de cualificación y temor al valor de los datos. La decisión de Hadoop se ha basado en una continua evolución del sistema aunque son muchos los foros que empiezan a desconfiar de la solución por las actualizaciones continuas también producen fallos de estabilidad.

Mi experiencia trabajando en el sector bancario me ha permitido ver que la problemática de la monetización digital es tangible y prioritaria existiendo un interés creciente por almacenar los datos de los clientes. Una de las implementaciones en las que he colaborado ha sido el acceso a conexión wifi en sucursales a través del perfil de redes sociales o el desarrollo de nuevas aplicaciones móviles que digitalicen la tarjeta de crédito para competir con ApplePay.

En el ámbito universitario la globalidad ha desmontado el concepto de “clase” como espacio físico, ahora el aprendizaje e interacción entre personas sin importar condición económica o ubicación geográfica es la única posible definición de aprendizaje universitario. Según los datos de un estudio de eMarketer y ResponseTap, un 80% de los profesionales de las agencias de consultoría y un 76% de los especialistas en relación con el cliente están usando Hadoop simplemente para almacenar datos. Los datos analizados son de 40000 usuarios lo que supone una problemática por el gran volumen y por la duplicidad en algunos casos entre diferentes redes o herramientas por ello es preciso de una gran cualificación dado que Hadoop no es capaz de dar “valor” a los datos. En mi experiencia en la Universidad tuve la suerte de formar parte de un gran equipo y ver que los datos son algo más que gráficos en un Excel.

Existe la visión para muchos de una universidad como una gran sala de videoconferencia (existen ya medios reales) pero en mi opinión creo que el factor humano y la relación es imprescindible y por ello no debemos dejarnos “ganar” por la batalla de los datos y crear una conciencia útil de uso de los nuevos medios.

Yo creo que la solución no se basa en digitalizar el contenido en pantallas táctiles sino en iniciativas como las de Prasad Ram, el jefe de Investigación y Desarrollo de Google India que dejó su carrera corporativa y creó su propia ONG, Goouru de aprendizaje complementado por Big Data. En AltSchool en San Francisco los alumnos cuando llegan a la escuela deben hacer un “check-in” en una App que controla la asistencia, las clases son grabadas y los alumnos dedican una parte del día a trabajar de forma independiente con playlists de actividades de acuerdo con sus objetivos personales.

Usando el Common Core Standards, diferentes escuelas o profesores podrán compartir entre ellos su conocimiento del progreso de un niño que se mueva de una clase a otra, o incluso de una escuela para otra.

Las redes sociales ya no son un simple lugar para compartir fotos ahora permiten la difusión del aprendizaje y “humanizar” la relación entre el alumno y el profesorado rompiendo por completo ese concepto de que un alumno es un nº y una nota en un examen final.

Evidentemente ese punto junto a la poca legislación existente lleva al **problema de la privacidad** Los usuarios están cada vez más preocupados por las cuestiones relacionadas con la privacidad de sus datos personales.

Legalmente, hay muchas regulaciones para proteger la privacidad especialmente cuando son niños. La ley COPPA en los EEUU manifiesta que cualquier operador de un servicio por internet tiene que “ofrecer información en su sitio web sobre qué datos sobre los niños estarán recopilados por el operador, y como serán usados”.

# Referencias

---

1. ZDNet.com, CBS Interactive, What is "Big Data?" 2013. :  
<http://www.zdnet.com/topic-big-data/>,
2. E. Plugge, P. Membrey & T. Hawkins, The Definitive Guide to MongoDB: The NoSQL Database for Cloud and Desktop Computing, Published Apress Media LLC, New York, 2010.
3. B. Hopkins, Beyond the Hype of Big Data. Disponible en:  
[http://www.cio.com/article/692724/Beyond\\_the\\_Hype\\_of\\_Big\\_Data](http://www.cio.com/article/692724/Beyond_the_Hype_of_Big_Data), 2011.
4. Big Data: oportunidades para las empresas en 2015  
<http://www.forbes.com.mx/big-data-oportunidades-para-las-empresas-en-2015/>
5. 9 estrategias definitivas de Big Data para el sector Seguros  
[http://cdn2.hubspot.net/hub/239039/file-1958725915-pdf/docs/PWD - Ebook - Seguros Big Data-  
\\_9 estrategias - FINAL v1.pdf](http://cdn2.hubspot.net/hub/239039/file-1958725915-pdf/docs/PWD_-_Ebook_-_Seguros_Big_Data_-_9_estrategias_-_FINAL_v1.pdf)
6. El científico de datos: una novedosa y necesaria profesión  
<http://noticias.universia.es/ciencia-nn-tt/noticia/2014/05/06/1095994/cientifico-datos-novedosa-necesaria-profesion.html>
7. La medicina del futuro pasa por big data  
<http://www.aunclidelastic.com/la-medicina-del-futuro-pasa-por-big-data/>
8. Las aplicaciones de big data en el deporte Think Big  
<http://blogthinkbig.com/big-data-deporte/>
9. Tecnología y ciencia con innovación entran  
<http://www.aunclidelastic.com/innovaciones-educativas-para-promover-la-vocacion-cientifico-tecnologica/>
10. Tendencias en Educación: la integración de las TIC de forma holística  
<http://www.aunclidelastic.com/tendencias-en-educacion-la-integracion-de-las-tic-de-forma-holistica/>
11. El año del MOOC:: Repaso a los avances en las TIC en Educación.  
<http://www.aunclidelastic.com/repaso-a-los-avances-tic-de-2012-ii/>
12. Telefónica Educación Digital  
<https://vimeo.com/telefonicaed>
13. SolidDQ Hadoop HDFS. Almacenamiento y consulta  
<http://blogs.solidq.com/es/big-data/hadoop-hdfs-almacenamiento-y-consulta-parte-1/>

14. Bases de datos NoSQL. Qué son y tipos que nos podemos encontrar  
<http://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>
15. TICBEAT 14 herramientas para una visualización de datos atractiva  
<http://www.ticbeat.com/bigdata/14-herramientas-para-una-visualizacion-de-datos-atractiva/>
16. SolidQ Hadoop HDFS. Almacenamiento y consulta  
<http://blogs.solidq.com/es/big-data/hadoop-hdfs-almacenamiento-y-consulta-parte-2/>
17. Happy Minds! Software: MAPREDUCE 2.0 ARQUITECTURA YARN  
<http://www.happyminds.es/apache-hadoop-introduccion-a-yarn/#sthash.bGSv0Hsd.dpbs>
18. ADQUISICIÓN DE DATOS Y GRABACIÓN: FLUME COMPUTING COMMUNITY CONSORTIUM  
<http://cra.org/ccc/>
19. Taller apache Cassandra Citius  
<http://eventos.citius.usc.es/bigdata/workshops/Cassandra.pdf>
20. VI Estudio Redes Sociales de IAB Spain  
[http://www.iabspain.net/wpcontent/uploads/downloads/2015/01/Estudio\\_Anual\\_Red\\_Sociales\\_2015.pdf](http://www.iabspain.net/wpcontent/uploads/downloads/2015/01/Estudio_Anual_Red_Sociales_2015.pdf)
21. Youtube EDU  
<https://www.youtube.com/channel/UC3yA8nDwraeOfnYfBWun83g>
22. Libro: Jakkob Nielsen . Usabilidad. Prioridad en diseño Web
23. Libro: Steve Krug . Don't make me think
24. Libro: Google Analytics de Ronan Chardonneau
25. Gartner Survey Reveals That 73 Percent of Organizations Have Invested or Plan to Invest in Big Data in the Next Two Years  
<http://www.gartner.com/newsroom/id/2848718>
26. El presidente del BBVA ve en Google, Amazon y Facebook a los grandes rivales de los bancos  
<http://www.elmundo.es/economia/2013/12/02/529ce33b61fd3da0228b456a.html>
27. 9 estrategias definitivas de Big Data para el sector Seguros  
[http://cdn2.hubspot.net/hub/239039/file-1958725915-pdf/docs/PWD\\_-\\_Ebook\\_-\\_Seguros\\_Big\\_Data-\\_9\\_estrategias\\_-FINAL\\_v1.pdf](http://cdn2.hubspot.net/hub/239039/file-1958725915-pdf/docs/PWD_-_Ebook_-_Seguros_Big_Data-_9_estrategias_-FINAL_v1.pdf)
28. Unversia: El científico de datos: una novedosa y necesaria profesión  
<http://noticias.universia.es/ciencia-nn-tt/noticia/2014/05/06/1095994/cientifico-datos-novedosa-necesaria-profesion.html>
29. La medicina del futuro pasa por big data  
<http://www.aunclidelastic.com/la-medicina-del-futuro-pasa-por-big-data/>
30. Las aplicaciones de big data en el deporte  
<http://blogthinkbig.com/big-data-deporte/>

31. Think Big Tecnología y ciencia con innovación entran  
<http://www.aunclidelastic.com/innovaciones-educativas-para-promover-la-vocacion-cientifico-tecnologica/>
32. NMC Horizon Report > 2014 Higher Education Edition  
<http://www.nmc.org/publication/nmc-horizon-report-2014-higher-education-edition/>
33. Sciencelab  
<http://www.science-lab.org/ueber-science-lab/>
34. **Kenneth Cukier**  
[https://www.ted.com/talks/kenneth\\_cukier\\_big\\_data\\_is\\_better\\_data?language=es](https://www.ted.com/talks/kenneth_cukier_big_data_is_better_data?language=es)
35. Telefónica Educación Digital  
<https://vimeo.com/telefonicaed>
36. Arcplan (2012). Big Data FAQs. A primer. Business Intelligence  
[www.arcplan.com/en/blog/2012/03./bigdata-faqs-a-primer/](http://www.arcplan.com/en/blog/2012/03./bigdata-faqs-a-primer/)
37. Armunanto, E. (2013). FAST: a machine to detect people who will soon break the law. Digital Journal. 02/04/2013.  
<http://www.digitaljournal.com/article/347078>
38. Arora, R. (2013). Location-based marketing and Big Data: aremarketers ready? The smart cube.  
<http://www.thesmartcube.com/insights/blog/business-strategy/locationbased-marketing-and-big-data-are-marketers-ready/>
39. BCSS (2013). Big data, small footprint? The Best ComputerScience Schools.  
<http://www.bestcomputerscienceschools.net/big-data/>
40. BOE (2013). Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.  
<http://boe.es/buscar/pdf/2013/BOE-A-2013-12887-consolidado.pdf>
41. Brannock, J. (2014). I knew you before I met you: how social media has changed the way we communicate. The Faculty Voice.  
<http://imerrill.umd.edu/facultyvoice1/?p=3171>
42. Brustein, J. (2014a). Consultant Andreas Weigend on Big Data refineries. Bloomberg Business Week. 06/03/2014. <http://www.businessweek.com/articles/2014-03-06/consultant-andreasweigend-on-big-data-refineries>
43. McKinsey (2013). Leveraging Big Data to optimize digital marketing. McKinsey & Company.  
[http://www.mckinsey.com/client\\_service/marketing\\_and\\_sales/latest\\_thinking/leveraging\\_big\\_data\\_to\\_optimize\\_digital\\_marketing](http://www.mckinsey.com/client_service/marketing_and_sales/latest_thinking/leveraging_big_data_to_optimize_digital_marketing)
44. 60. Mello, U., Treinish, L. (2012). Precision agriculture, using predictiveweather analytics to feed future generations. IBM Research.  
[http://www.research.ibm.com/articles/precision\\_agriculture.shtml](http://www.research.ibm.com/articles/precision_agriculture.shtml)
45. 61. Mills, M. P. (2013). The cloud begins with coal. Big Data, big networks, big infrastructure, and big power. An overview of the electricity used by the global ecosystem. Digital Power Group.  
<http://www.tech-pundit.com/wp-content/uploads/2013/07/>
46. Naughton, J. (2014). We're all being mined for data – but who are the real winners? The Guardian.  
<http://www.theguardian.com/technology/2014/jun/08/big-data-mined-realwinners-nsa-gchq-surveillance>
47. Nielson, B. (2013). MOOC analytics: what corporate training can learn from Big Data. Your Training Edge. 24/06/2013.  
<http://www.yourtrainingedge.com/mooc-analytics-whatcorporate-training-can-learn-from-big-data/>

# GLOSARIO Big Data

---

## A

API - interfaz de programación de aplicaciones que define la manera de trabajar con distintos componentes de programación.

Arquitectura maestro-esclavo - arquitectura de red de servidores jerarquizada donde el nodo maestro lleva el peso del control mientras que los esclavos realizan el trabajo de procesamiento.

Apache Flume; Es un marco para aportar datos a Hadoop. Los agentes residen en toda la infraestructura de TI como servidores web, servidores de aplicaciones y dispositivos móviles, para recoger esos datos e integrarlos en Hadoop.

Apache Hive: Es un sistema de almacenamiento de datos sobre Hadoop al os que añade metadata para facilitar su manejo, creando lo que se llama un almacén, desarrollado originalmente por Facebook. Permite a los usuarios escribir consultas SQL en un lenguaje denominado HiveQL, que luego se convierte en MapReduce.

Apache ZooKeeper Provee un servicio de configuración centralizada y registro de nombres de código abierto para grandes sistemas distribuidos.

Avro Es un sistema de serialización de datos optimizado para Hadoop/MapReduce. Tiene la ventaja de ser compacto, flexible y admitir varios lenguajes de programación.

## B

Base de datos: Contenedor informático que permite reunir la información relativa a una actividad.

Big Data: Conjunto de datos tan grande que es difícil trabajar con los medios habituales (bases de datos).

BANNER: Es un anuncio publicitario que se encuentra enlazado en un sitio web.

## C

Clúster - conjunto de ordenadores interconectados que actúan como si fueran uno solo.

Cassandra: Tienda de datos NoSQL, Apache Cassandra, es un sistema de gestión distribuido de bases de datos desarrollado por Facebook para potenciar su función de búsqueda en la bandeja de entrada.

Caché Web El caché es un componente que almacena datos y documentos web como imágenes, paginas, audios, etc;...

Cookie: es un archivo de texto de pequeño tamaño que se descarga de forma automática en algunos sitios de internet al navegar y contiene información sobre la actividad realizada por el usuario en esa web.

## D

Data warehouse - es una colección de datos orientada a un ámbito empresarial u organizativo y almacenada en un sistema no volátil e integrado que ayuda a la toma de decisiones.

Dashboard - o tablero de instrumentos es una interfaz desde la que el usuario puede controlar y administrar una aplicación.

Distribuido - la computación distribuida es un modelo de computación en la que se usan una serie de ordenadores organizados en clústeres.

## E

Escalabilidad - es una propiedad de un sistema que indica su capacidad de reacción y de adaptación a los cambios de envergadura.

ETL - proceso de transformación de datos realizado para extraer datos de una fuente y almacenarlos en una base de datos o data warehouse.

## F

Framework - conjunto de conceptos y tecnologías que sirven de base para el desarrollo de aplicaciones. Acostumbra a incluir bibliotecas de software, lenguajes, soportes a aplicaciones de desarrollo...

## H

Hadoop Apache: software de código abierto para aplicaciones intensivas de datos distribuidos a través de MapReduce, Hadoop distribuye los datos en grandes piezas sobre una serie de nodos que se ejecutan en hardware de consumo.

HBase Escrito en Java es una base de datos no relacional distribuida en columnas, diseñada para ejecutarse en la parte superior de Hadoop Distributed Filesystem (HDFS).

HDFS (Hadoop Distributed File System) Es la capa de almacenamiento de Hadoop, es un sistema de fichero distribuido escrito en java, escalable, tolerante a fallos y compatible con MapReduce.

## I

IDE - Integrated Development Environment y consiste en una aplicación que agrupa un conjunto de herramientas -como editor de texto, gestor de proyectos, compilador o debugger- para facilitar el desarrollo de aplicaciones.

## L

Log - es un fichero que archiva un conjunto de entradas que informan de los distintos eventos que se producen en un ordenador.

## M

Modelo relacional - es un modelo de datos basado en la lógica de predicados y en la teoría de conjuntos. Está compuesto fundamentalmente por tablas y se basa en el uso de las relaciones entre estas.



Mahout Es una librería de algoritmos de machine learning. En ella están los algoritmos de minería de datos más populares para llevar a cabo la agrupación, pruebas de regresión y modelos estadísticos implementados usando MapReduce para que puedan ejecutarse sobre Hadoop

MapReduce Patrón de arquitectura que permite realizar cálculos en paralelo y, por tanto, perfectamente adecuado para el tratamiento de bases de datos de gran tamaño.

## N

NoSQL - clase de sistemas de gestión de bases de datos que ofrecen una alternativa al modelo relacional tradicional.

## O

Open-source - hace referencia al código distribuido y desarrollado libremente, dando acceso al código fuente del proyecto.

## P

Phishing - ataque informático que consiste en suplantar la identidad de una persona o una entidad para engañar al atacado.

Repositorio - servidor centralizado donde se almacena y mantiene información digital para poder ser accedidos remotamente.

## S

Serialización/deserialización de datos - codificación o decodificación de un objeto en un medio con el fin de transmitirlo en forma de series de bytes.

SQL - lenguaje de consultas estructuradas para bases de datos relacionales.

Streaming - distribución de datos de manera constante en forma de flujo continuo.

*T*

Timestamp - secuencia de caracteres que denotan una fecha y hora.

*V*

Virtualización - emular un entorno físico mediante software dentro de otro entorno físico real.

*W*

Workflow - flujo de trabajo que define como tienen que ejecutarse y comunicarse entre ellas diversas aplicaciones o ejecuciones.



# Acrónimos del Marketing Digital

---

CPM	<i>Coste Por Mil Impresiones</i>
CTR	<i>Click Through Rate/Tasa de Clic</i>
CPC	<i>Coste Por Clic</i>
CPA	<i>Coste Por Adquisición/Venta</i>
CPL	<i>Coste Por Lead</i>
eCPM	<i>CPM efectivo</i>
DSP	<i>Demand Side Platform</i>
SSP	<i>Sale Side Platform</i>
RTB	<i>Real Time Bidding</i>
DMP	<i>Data Management Platform</i>
API	<i>Application Program Interface</i>
SEO	<i>Search Engine Optimization</i>
SEM	<i>Search Engine Marketing</i>
B2B	<i>Business to Business</i>
B2C	<i>Business to Consumer</i>
BR	<i>Bounce Rate/Tasa de Rebote</i>
CMS	<i>Content Management System</i>
APP	<i>Application/Aplicación móvil</i>
CTA	<i>Call To Action/Llamada a la acción</i>
CRM	<i>Customer Relation Management</i>
KPI	<i>Key Performance Indicator</i>
ROI	<i>Return On Investment</i>
SMO	<i>Social Media Optimization</i>
TT	<i>Trending Topic</i>
URL	<i>Unique Resource Locator</i>

<b>UX/CX</b>	<i>User Experience/Customer Experience</i>
<b>HTML</b>	<i>HyperText Markup Language/Lenguaje de marcas de hipertexto</i>
<b>LTV</b>	<i>Life Time Value</i>
<b>PPC</b>	<i>Pay Per Click</i>
<b>UGC</b>	<i>User Generated Content/Contenido Generado por los Usuarios</i>

---

# Hadoop: instalación y configuración

---

## Fase previa

- 1) Instalar Java JDK ejecutando los siguientes comandos:

```
sudo apt-get update
```

```
sudo apt-get install default-jdk
```

- 2) Configurar el acceso SSH para la gestión de los nodos.

```
sudo apt-get update
```

```
sudo apt-get install ssh
```

- 3) Configurar el acceso ssh del usuario y generar una clave SSH:

```
su - hduser
```

```
ssh-keygen -t rsa -P ""
```

- 4) Habilitar el acceso SSH a la máquina local:

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

- 5) Probar que se ha configurado correctamente el acceso SSH:

```
ssh localhost
```

- 6) Crear grupo de usuarios al que pertenecerá el usuario de Hadoop

```
sudo addgroup hadoop
```

```
sudo adduser --ingroup hadoop hduser
```

- 7) Añadir el usuario creado al grupo sudo para que pueda ejecutar aplicaciones en modo administrador.

```
sudo adduser hduser sudo
```

## Instalar Hadoop Common

- 1) Entrar en (<http://hadoop.apache.org/>) y descargar Apache Hadoop Common.

```
tar -xzf hadoop-2.6.0.tar.gz
```

- 2) Mover los archivos a su ubicación /usr/local:

```
sudo mv hadoop-2.6.0/ /usr/local
```

- 3) Crear un enlace simbólico a la carpeta que contiene los ficheros:

```
cd /usr/local/
```

```
sudo ln -s hadoop-2.6.0/ hadoop
```

- 4) Establecer la variable de entorno HADOOP\_HOME:

```
export HADOOP_HOME=/usr/local/hadoop
```

- 5) Cambiar el usuario propietario de los archivos de Hadoop Common instalados.

```
sudo chown -R hduser:hadoop /usr/local/hadoop-2.6.0
```

- 6) Configurar Hadoop como un clúster de un sólo nodo.

```
/usr/local/hadoop/bin/hadoop version
```

- 7) Configurar el directorio donde Hadoop va a guardar los archivos de datos

```
mkdir /usr/local/hadoop/data
```

- 8) Asignar *hduser* propietario.

```
sudo chown hduser:hadoop /usr/local/hadoop/data
```

- 9) Editar el fichero `$(HADOOP_HOME)/etc/hadoop/hadoop-env.sh`:

```
# set to the root of your Java installation
```

```
export JAVA_HOME=/usr/java/latest
```

```
# Assuming your installation directory is /usr/local/hadoop
```

```
export HADOOP_PREFIX=/usr/local/hadoop
```

- 10) Editar el fichero `$(HADOOP_HOME)/etc/hadoop/core-site.xml`:

Insertar las siguientes líneas dentro de las etiquetas `<configuration>` `</configuration>`.

```
<property>
```

```
<name>fs.defaultFS</name>
```

```
<value>hdfs://localhost:9000</value>
```

```
<description>The name of the default file system. A URI whose scheme and authority determine the FileSystem implementation.</description>
```

```
</property>
```

- 11) Editar el fichero `$(HADOOP_HOME)/etc/hadoop/hdfs-site.xml`:

Insertar las siguientes líneas entre las etiquetas `<configuration>` `</configuration>`:

```
<property>
```

```
<name>dfs.replication</name>
```

```
<value>1</value>
```

```
<description>Default block replication. The actual number of replications can be specified when the file is created. The default is used if replication is not specified in create time.</description>
```

```
</property>
```

12) Editar el fichero \$HADOOP\_HOME/etc/hadoop/mapred-site.xml:

Editar el fichero e insertar las siguientes líneas dentro de las etiquetas <configuration> </configuration>.

```
<property>
```

```
<name>mapreduce.framework.name</name>
```

```
<value>yarn</value>
```

```
</property>
```

```
<property>
```

```
<name>mapred.job.tracker</name>
```

```
<value>localhost:54311</value>
```

```
<description>The host and port that the MapReduce job tracker runs at. If "local", then jobs are run in-process as a single map and reduce task.</description>
```

```
</property>
```

13) Editar el fichero \$HADOOP\_HOME/etc/hadoop/yarn-site.xml:

Insertar las siguientes líneas entre las etiquetas <configuration> </configuration>:

```
<property>
```

```
<name>yarn.nodemanager.aux-services</name>
```

```
<value>mapreduce_shuffle</value>
```

```
</property>
```

14) Actualizar \$HOME/.bashrc

Añadir las siguientes líneas al final del fichero \$HOME/.bashrc para el usuario *hduser*.

```
# Set Hadoop-related environment variables
```

```
export HADOOP_HOME=/usr/local/hadoop
```

```
# Set JAVA_HOME (we will configure JAVA_HOME for Hadoop later on)
```

```
export JAVA_HOME=/usr/lib/jvm/default-java
```

```
# Some convenient aliases and functions for running Hadoop-related commands
```



```

unalias fs &> /dev/null

alias fs="hadoop fs"

unalias hls &> /dev/null

alias hls="fs -ls"

# If you have LZOP compression enabled in your Hadoop cluster and
# compress job outputs with LZOP
# Conveniently inspect an LZOP compressed file from the command
# line; run via:
#
# $ lzohead /hdfs/path/to/lzop/compressed/file.lzo
#
# Requires installed 'lzop' command.
#
lzohead () {
    hadoop fs -cat $1 | lzop -dc | head -1000 | less
}

# Add Hadoop bin/ directory to PATH
export PATH=$PATH:$HADOOP_HOME/bin

```

## Formatear el sistema de archivos HDFS

### 1) Formatear el sistema de archivos del cluster

```
/usr/local/hadoop/bin/hadoop namenode -format
```

### 2) Arrancar el Clúster

```
/usr/local/hadoop/sbin/start-all.sh
```

### 3) Comprobar que procesos de Hadoop que se están ejecutando.

```
jps
```

Esto devolverá un mensaje por pantalla:

```
xxxx DataNode
```

```
xxxx NodeManager
```

```
xxxx ResourceManager
```

xxxx Jps

xxxx SecondaryNameNode

xxxx NameNode

- 4) Comprobar si Hadoop está escuchando en los puertos configurados.

`sudo netstat -plten | grep java`

## Subir ficheros a HDFS

- 1) Crear un directorio HDFS

`hdfs dfs -mkdir /user`

`hdfs dfs -mkdir /user/hduser`

- 2) Copiar archivos en el sistema de ficheros distribuido

`hdfs dfs -put <carpeta_local_origen> <carpeta_hdfs_destino>`

- 3) Borrar un archivo del sistema de ficheros distribuido

`hdfs dfs -rm <carpeta_hdfs>/<nombre_fichero>`

- 4) Parar el Cluster

`/usr/local/hadoop/sbin/stop-all.sh`

- 5) Acceso a la interfaz WebUI para seguimiento de los Jobs.

<http://localhost:8088/cluster/nodes>

Nodes of the cluster

Cluster Metrics												
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	1	0	0	0	0

Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail
/default-rack	RUNNING	localhost:38490	localhost:8042	24-Feb-2014 00:17:38		0	0 B	8 GB

Showing 1 to 1 of 1 entries



# Formación y certificados

---

BT Program in Global Services for Santander Bank

## BRITISH TELECOM SALES



*Junio 2015-Actualmente*

British Telecom es un operador de comunicaciones británico con más de 1000 trabajadores en España que da servicio de red a grandes empresas, pymes y organismos públicos, funcionando, a su vez, como consultora.

Actividad profesional en el area de Sales en la gestion de la cuenta del Banco Santande. Entre mis principales funciones destaca:

- Realizacion de ofertas de servicios profesionales para redes de telecomunicaciones.
- Desarrollo de actividad consultora en el analisis de datos financieros de la cuenta.
- Estudio de viabilidad de propuestas de servicios de telecomunicación.
- Gestion del cliente para la realizacion de nuevas propuest a.

## Gabinete de comunicación – Rectorado de la Universidad Autonoma de Madrid.

Ejecución del Plan Estratégico de Comunicación e Imagen Corporativa de la UAM bajo las ordenes del Delegado del Rector para Comunicación e Imagen Corporativa.

- Desarrollo de políticas de comunicación en Redes Sociales institucionales (Community Management).
- Colaboración en el proyecto “Google Indoor Maps UAM”.
- Desarrollo de estrategias SEO para el fortalecimiento internacional de la marca UAM.
- Consultoria en la ejecucion de estudios sobre de los diferentes estamentos de la Universidad.
- Asesoramiento a equipos de trabajo e investigacion de la Universidad Autonoma de Madrid en el lanzamiento y consolidacion de su actividad digital y difusion de contenidos.
- Colaboración con Oracle en el desarrollo de la nueva página home UAM.

## CERTIFICACIONES



### ➤ **MARKETING ONLINE. ESTRATEGIAS Y HERRAMIENTAS PARA EL ANALISIS Y ÉXITO.**

**Certificado acreditativo de Google e IAB.**

Panorama digital. | Tecnologías de medición. SEO-SEM. Analítica web y modelos de negocio. Display Advertising. | Mobile marketing. | Redes sociales. | Plan de marketing. | E-commerce-



### ➤ **EMPRENDIMIENTO.COMERCIO ELECTRÓNICO. IMPLEMENTACION Y TRANSACCIONES.**

**Certificado acreditativo de la Escuela de Organización Industrial**

Mobile commerce. | Digitalización de planes de negocio. | Logística y distribución. | Legislación y política de datos de datos en e-commerce. Publicidad digital. | Redes sociales en el e-commerce.



### ➤ **ANALÍTICA WEB DE DATOS EN MARKETING DIGITAL. POSICIONAMIENTO SEO/SEM**

**Certificado acreditativo de la Escuela de Organización Industrial**

Análisis de datos para empresas. | Análisis de las estrategias de la competencia. | Análisis del posicionamiento de marca en buscadores. | Desarrollo de estrategias para mejorar la experiencia online de los clientes. Creación de campañas mediante Google Adwords.



### ➤ **MARKETING EN REDES SOCIALES Y MÓVILES.**

**Certificado acreditativo de la Universidad Autónoma de Madrid.**

Investigación de Mercados en Redes Sociales. | Investigación de Reputación Online. Investigación de Tendencias: Coolgunning Digital. | Investigación de Resultados: Analítica Web.

---

## CURSOS

### ➤ LIDERAZGO EN ORGANIZACIONES.

Curso impartido por la *Facultad de Psicología de la Universidad Autónoma de Madrid*.

Exponer y desarrollar el conjunto de habilidades personales, conductuales y situaciones que acompañan la emergencia y la eficacia de los líderes dentro de sus organizaciones.

### ➤ DESARROLLO DE COMPETENCIAS PROFESIONALES: NEGOCIACIÓN.

Curso impartido por *Instituto de Ingeniería y Conocimiento*.

Herramientas y métodos para mejorar su eficacia en sus negociaciones. Fundamento de los estilos de negociación. Promover estrategias positivas en negociaciones win to win, Análisis y características de las distintas fases de una negociación.

### ➤ CURSO DE ORATORIA Y HABILIDADES PARA HABLAR EN PÚBLICO

Curso impartido por el catedrático de la *Facultad de Psicología*, Don José Santecreu Mas.

Formación integral de habilidades para la exposición comercial o técnica ante públicos diversos. Aprendizaje de técnicas para captar la atención y manejar audiencias.

### ➤ TICS EN MÉDICINA: CASA DOMOTICA Y TELEMEDICINA EN EL FUTURO.

Curso impartido por la *Universidad Politécnica de Madrid*.

Prestación de servicios a distancia o mediante monitorización en situaciones de enfermedad crónica o en situaciones de dependencia.

### ➤ DESARROLLO EN COMPETENCIAS “Flexibilidad y adaptación”

Curso impartido por el Instituto de *Ingeniería y Conocimiento de la UAM*.

Habilidades para gestionar la dedicación laboral de un modo más efectivo, según las necesidades de los proyectos y los futuros cambios que se puedan producir tomando de forma estratégica las mejores decisiones.

# Presupuesto solución Big Data redes sociales UAM

Ejecución Material	NRC	RC	Total solución 1 año
Compra de ordenador personal iMac 2.8GHz Processor 1TB Storage	2.000 €	0 €	2.000 €
Compra impresora HP LaserJet Pro MFP M225dw	239 €	0 €	239 €
Material de oficina	250 €	0 €	250 €
Router Teldat M alquiler y gestión	465 €	0 €	465 €
Servidores	0 €	400 €	4.800 €
<b>Total de ejecución material</b>			<b>7.754 €</b>

Licencias	NRC	RC	Total solución 1 año
Licencia Office 365 2016 for Mac	87 €	0 €	87 €
Almacenamiento 100Gb Google Drive (1 año)	24 €	0 €	24 €
Licencia de software para almacenamiento de datos	0 €	279 €	3.348 €
Licencia Wordpress premium	99 €	0 €	99 €
Plan premium Wix servidor webb	0 €	5 €	60 €
<b>Total de licencias</b>			<b>3.618 €</b>

Herramientas analítica	NRC	RC	Total solución 1 año
SocialBro PlanPro	0 €	39 €	468 €
Google Analytics Premium	0 €	20 €	240 €
Prezi plan premium	0 €	20 €	240 €
Woorank Pro Plan	0 €	50 €	600 €
Hootsuite	0 €	18 €	216 €
<b>Total de herramientas de analítica</b>			<b>1.764 €</b>

Gastos generales	NRC	RC	Total solución 1 año
16 % sobre Ejecución Material	1.240 €	0 €	1.240 €

Beneficio Industrial	NRC	RC	Total solución 1 año
6 % sobre Ejecución Material	465 €	0 €	465 €

Honorarios Proyecto	NRC	RC	Total solución 1 año
740 horas a 15 € / hora	11.100 €	0 €	11.100 €

<b>Material fungible</b>	<b>NRC</b>	<b>RC</b>	<b>Total solución 1 año</b>
<b>Gastos de impresión</b>	60 €	0 €	60 €
<b>Encuadernación</b>	200 €	0 €	200 €
<b>Total de ejecución material</b>			260 €
<b>Subtotal del presupuesto</b>	<b>NRC</b>	<b>RC</b>	<b>Total solución 1 año</b>
Subtotal Presupuesto	26.201 €	0 €	26.201 €
<b>I.V.A. aplicable</b>	<b>NRC</b>	<b>RC</b>	<b>Total solución 1 año</b>
21% Subtotal Presupuesto	5.502 €	0 €	5.502 €
<b>Total presupuesto</b>		31.703 €	

Madrid, Diciembre de 2015

El Ingeniero Jefe de Proyecto

Fdo.: Ismael Herrera De los Ríos

Ingeniero de Telecomunicación



# **PLIEGO DE CONDICIONES**

---

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de *"Estrategias, herramientas y métricas de posicionamiento de Big Data en laboratorios de investigación universitarios en redes sociales"*. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

## **Condiciones generales**

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez

la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

## **Condiciones particulares**

---

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime

oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.