

# Big Data: El valor de los Datos

Estado actual y tendencias del Big Data como nuevo activo en la economía europea.

Instituto Tecnológico de Informática (ITI)



Cofinanciado por:



---

**Big Data: El valor de los Datos**

**Estado actual y tendencias del Big Data como nuevo activo en la economía europea.**

María Antolín Fernández

Francisco Ricau González

Daniel Sáez Domingo

*Han colaborado en el monográfico:* Francisco Sánchez Cid y Raúl Hussein Galindo.

Editado por Colegio Oficial de Ingenieros en Informática de la Comunidad Valenciana (COIICV)

CIF: V-97046189

Datos de contacto:

Av. Barón de Carcer 48, 3ºO. 46001 – Valencia

963622994 – secretaria@coiicv.org

[www.coiicv.org](http://www.coiicv.org)

ISBN: 978-84-606-7474-0

**Publicación gratuita (prohibida su venta)**

Primera edición: Abril 2015

Todos los derechos reservados

Se autoriza la reproducción total o parcial de este informe con fines educativos, investigadores, divulgativos y no comerciales citando la fuente. La reproducción para otros fines queda expresamente prohibida sin la autorización del Instituto Tecnológico de Informática y del COIICV.



# Índice

<b>0</b> p.4	<b>Presentación ITI</b>		
<b>0</b> p.5	<b>Presentación COICV</b>		
<b>0</b> p.7	<b>Prólogos</b>		
<b>0</b> p.11	<b>Resumen ejecutivo</b>		
<b>1</b> p.13	<b>Introducción</b>		
	1.1 Justificación y objetivos del estudio	13	
	1.2 Contexto del Big Data	14	
	1.2.1 ¿Qué es Big Data?	14	
	1.2.2 ¿De dónde provienen todos estos datos?	15	
	1.2.3 ¿Qué ventajas supone adoptar Big Data?	17	
	1.2.4 ¿Qué grado de adopción empresarial tiene el Big Data?	18	
	1.2.5 ¿Qué impacto supondrá el Big Data?	20	
	1.2.6 Big Data versus Open Data	20	
<b>2</b> p.22	<b>Cadena de valor</b>		
	2.1 Cadena de valor y principales "stakeholders"	22	
	2.2 Áreas científico-tecnológicas prioritarias	26	
	2.3 Ecosistema de tecnologías existentes	31	
	2.3.1 Tecnología base	32	
	2.3.2 Infraestructura	32	
	2.3.3 Análisis	34	
	2.3.4 Aplicaciones	35	
<b>3</b> p.36	<b>Agentes e iniciativas relevantes</b>		
	3.1 Proveedores	36	
	3.2 Plataformas tecnológicas y entidades asociadas	39	
	3.3 Proyectos I+D+I	45	
<b>4</b> p.50	<b>Dominios de aplicación</b>		
	4.1 Sectores principales e impactos esperados	50	
	4.2 Barreras de entrada	67	
<b>5</b> p.71	<b>Visión y contribución de ITI al Big data</b>		
	5.1 Descripción general de ITI	71	
	5.2 Capacidades y experiencias relacionadas	73	
<b>6</b> p.76	<b>Bibliografía</b>		

# Índice de figuras y tablas

**F1**

p.14

## Figura 1

Dimensiones de Big Data (Fuente: IBM [1])

**F2**

p.17

## Figura 2

Tipos de datos de Big Data (Fuente: Sunil Soares, Information Asset, LLC, 2012)

**F3**

p.19

## Figura 3

Fases del patrón de la adopción de Big Data (Fuente: IBM [1])

**F4**

p.19

## Figura 4

Hiperciclo para Tecnologías Emergentes 2014 (Fuente: Gartner)

**F5**

p.21

## Figura 5

Coste-beneficio de la apertura de los datos para sus propietarios (Fuente: WISE Institute, [3])

**F6**

p.22

## Figura 6

The data value chain (Fuente: IT Profesional Magazine, IEEE Computer Society [4])

**F7**

p.23

## Figura 7

Cadena de valor Big Data para el caso de uso de datos de tráfico (Fuente: Innosight, consultora estratégica y de innovación, 2013)

**F8**

p.25

## Figura 8

Ecosistema de Big Data a lo largo de la cadena de valor (Fuente: BDVA [6])

**F9**

p.25

## Figura 9

Actores principales del ecosistema de Big Data (Fuente: BDVA [6])

**F10**

p.31

## Figura 10

Tecnologías del Ecosistema de Big Data (Fuente: Elaboración Propia)

**F11**

p.36

## Figura 11

Worldwide Big Data Technology and Services 2014–2018 Forecast (Fuente: IDC)

**F12**

p.37

## Figura 12

Big Data Vendor Revenue and Market Forecast 2013-2017 (Fuente: Wikibon 2014)

**F13**

p.40

## Figura 13

Lista de miembros fundadores de la Big Data Value Association

**F14**

p.43

## Figura 14

Smart Data Innovation Lab Structure

**F15**

p.44

## Figura 15

TeraLab Infrastructure

**F16**

p.45

## Figura 16

SURFSARA Infrastructure

**F17**

p.51

## Figura 17

Incremento del PIB por sectores en Europa y en 2020 debido a soluciones basadas en datos (Fuente: WISE Institute, [3])

**F18**

p.52

## Figura 18

Potencial valor financiero generado por Big Data en diferentes sectores (Fuente: McKinsey [7])

**F19**

p.53

## Figura 19

Posicionamiento de los sectores para obtener beneficio del Big Data (Fuente: Fuente: McKinsey [7])

**F20**

p.64

## Figura 20

Tecnologías facilitadoras de la Ciudad Inteligente (Fuente: ITI [14])

**F21**

p.67

## Figura 21

Obstáculos clave para la adopción de Big data (Fuente: IBM [1])

**F22**

p.68

## Figura 22

Barreras en la adopción de Big Data (Fuente: OBS)

**F23**

p.72

## Figura 23

Áreas de especialización TIC del ITI

**T1**

p.38

## Tabla 1

2013 Worldwide Big Data Revenue by Vendor (\$US millions) (Fuente: Wikibon)

**T2**

p.47

## Tabla 2

Proyectos de I+D+i internacionales relacionados con Big Data (Fuente: Elaboración propia)

# PRESENTACIÓN ITI

**José Manuel Bernabéu Auban**

**DIRECTOR CIENTÍFICO**

**INSTITUTO TECNOLÓGICO DE INFORMÁTICA (ITI)**



Tengo el placer de presentar nuestro segundo trabajo monográfico realizado en colaboración con el Colegio Oficial de Ingenieros en Informática de la Comunidad Valenciana (COIICV). En esta ocasión, no podíamos menos que abordar la temática de la extracción de información de las grandes cantidades de datos, que hoy en día es posible obtener y almacenar, y cuyas técnicas se enmarcan dentro del término genérico de Big Data.

La información siempre ha sido el factor clave en la toma de decisiones en todos los ámbitos de la actividad humana. Sin embargo, la adquisición de información de calidad, en base a la cual poder tomar decisiones, jamás ha sido una tarea fácil. Como ejemplo sirva el de la actividad científica, centrada en buena parte en la obtención de información utilizable de forma sistemática.

El despegue de las TIC, con su capacidad para procesar grandes cantidades de información ha facilitado enormemente su análisis, conduciendo a la automatización masiva en la toma de decisiones en todos los ámbitos, desde los sistemas de gestión empresarial, a los de producción y fabricación.

Desde bien pronto no pasó desapercibida la capacidad de utilizar la información estructurada disponible para derivar de ella nueva información de gran valor, mostrando los patrones escondidos en la información explícitamente manejada para otros fines.

Hoy en día, gracias a la evolución de los sistemas de información, muchas de las actividades que realizamos están mediadas por alguna aplicación informática, capaz de generar, capturar y almacenar grandes cantidades de datos que antes no estaban disponibles en soporte digital. Dichos datos ofrecen la oportunidad de obtener información valiosa.

Algunas estimaciones calculan que la cantidad de datos almacenados se sitúa en el entorno de varios zetabytes ( $10^{21}$  bytes), cantidad que va aumentando exponencialmente de forma imparable. Cada vez son más las vías de interacción interpersonal mediatizadas por algún dispositivo capaz de capturarlas y almacenar sus contenidos. Cada vez hay más dispositivos, sensores de todo tipo desplegados en plantas industriales, hogares, nuestras calles e, incluso, nosotros mismos. Mientras, empresas de todos los sectores van tomando conciencia del valor potencial de los datos que manejan, y que hasta la fecha simplemente eran ignorados.

En definitiva, el mundo se convierte en un enorme laboratorio, proporcionando un flujo ingente de datos y la promesa de obtener información potencialmente valiosa en muchos ámbitos de nuestra sociedad. Ser capaces de extraer la información que de seguro se esconde en estos datos plantea grandes retos, cuya solución será la clave de la competitividad de nuestras empresas en el futuro cercano.

**Valencia, abril de 2015**

# PRESENTACIÓN COIICV

**Juan Pablo Peñarrubia Carrión**

**PRESIDENTE**

**Colegio Oficial de Ingenieros en Informática  
de la Comunidad Valenciana (COIICV)**



Las técnicas de obtención y tratamiento masivo de información se han consolidado como una de las principales tendencias y áreas de actividad e innovación aplicada de la ingeniería informática en los últimos años.

Este ámbito que ha dado denominarse Big Data ha sido en gran medida impulsado por la evolución e intensificación de otras áreas de aplicación de las tecnologías de la información, entre las que cabría destacar: Redes sociales y social media; Computación móvil; Monitorización de seguridad y seguridad proactiva; Smart cities; Gestión integral de las relaciones con el cliente (CRM integral); Sistematización del análisis financiero, de costes y de operaciones; gestión informática integral de infraestructuras y edificios basada en la sensorización; Informática industrial (especialmente en producción y robotización); Márketing digital; Ocio digital; Internet de las cosas...

Big Data, junto con el resto de tendencias informáticas con mayor impacto y evolución en estos años, no solo está afectando a la actividad de los profesionales y departamentos de informática, sino que en muchos casos está generando una transformación del conjunto de la organización.

Big Data ha significado en muchas organizaciones el concepto cualitativo que necesitaban para comprender el potencial del tratamiento informático de datos. En muchos casos para frustración de los CIOs y los departamentos de informática, incomprensidos como voces en el desierto. Este potencial, si bien siempre ha estado ahí, permite actualmente una correlación de áreas de actividad de la organización y una concepción integral, que abren infinitas posibilidades: mejora de la toma de decisiones, productividad, rentabilidad... y en definitiva competitividad y mejor estrategia de la organización.

Podríamos decir sin miedo a equivocarnos que con Big Data muchos empresarios, directivos y organizaciones han tenido por fin la epifanía, la visión, de uno de los potenciales básicos de la informática, hasta el momento infraaprovechado.

En el sector empresarial y especialmente el industrial, Big Data es una pieza básica de lo que se está dando en llamar la economía digital y el proceso de digitalización. Conceptos que los empresarios han hecho suyos como nueva concepción de la empresa. Una concepción en la que la informática pasa de ser una parte más de la organización, a ser materia prima esencial de la empresa y su evolución.

Verdaderamente Big Data abre infinitas posibilidades para la automatización, la ayuda a la decisión y la mejora continua de productos y servicios. Y en definitiva del negocio.

Como recoge el título de esta monografía, se trata del valor de los datos como un nuevo activo en la economía de la empresa.

**Valencia, abril de 2015**

# PRÓLOGO

**Lucía Álvarez**

**DIRECTORA DE ANALYTICS EN IBM ESPAÑA, PORTUGAL, GRECIA E ISRAEL**



Es indudable que nuestro día a día está cambiando. Junto con la proliferación de sensores, la tecnología que permite monitorizar lo que ocurre a nuestro alrededor y que nos informa de ello a través de nuestros dispositivos móviles se extiende de manera firme. Ahora podemos saber a qué hora pasará un autobús u otro y elegir según nos convenga. También podemos ver cuánta energía estamos gastando en casa y decidir apagar un electrodoméstico para reducir el gasto si eso es lo que queremos.

Otra de las novedades tecnológicas importantes de los últimos años es la extensión de la participación de la población en las redes sociales. Podemos comunicarnos con personas de cualquier parte del planeta de manera inmediata, sencilla y barata.

Un tercer factor de cambio es la nueva capacidad de almacenar, acceder y analizar gran cantidad de información, incluso en tiempo real.

Aparte de las ventajas que todo esto supone para nosotros como individuos en nuestro día a día, la suma de los tres factores mencionados abre un nuevo mundo a las empresas, que pueden tomar decisiones basadas en una mayor cantidad de información y centrar su estrategia en lo que verdaderamente le interesa a su cliente, elevando sus posibilidades de éxito y de fidelización.

Pero, además de los objetivos centrados en el cliente, las primeras aplicaciones de Big Data permiten abordar otros objetivos más funcionales, como, por ejemplo, la optimización operativa, la gestión financiera y de riesgos, la colaboración de los empleados y la habilitación de nuevos modelos de negocio.

De lo que no cabe duda es que la tendencia de Big Data va en aumento, como demuestra el hecho de que el Institute for Business Value de IBM publique que en 2014 había 10.000 millones de dispositivos móviles y que se proyecten 30.000 millones para 2020 y más de 100.000 millones para 2050. Según el estudio "Big Data @ Work Study" del mismo Institute of Business Value de IBM, realizado en colaboración con la Saïd Business School de la Universidad de Oxford, de las más de 1.000 empresas encuestadas en más de 130 países un 63% afirma que el uso de las tecnologías analíticas está creando una ventaja competitiva para sus organizaciones, lo cual supone un 37% más de las que habían hecho esta misma afirmación en el estudio realizado dos años antes.

Los pasos estratégicos fundamentales que las empresas han de seguir para adoptar iniciativas de Big Data con éxito serían:

- **Centrarse en los clientes:** Para muchas organizaciones, el gran valor asociado a Big Data se encuentra en el área de los clientes. Utilizando esta tecnología pueden entenderse las preferencias de los clientes y anticipar comportamientos futuros, lo que permite a las compañías mejorar significativamente la satisfacción por la compra de sus productos y aumentar la fidelidad hacia su marca.
- **Empezar con la información existente:** Para conseguir resultados en Big Data las compañías están llevando a cabo una aproximación pragmática empezando por la información interna ya existente, como la de las transacciones en el punto de venta o las llamadas en los servicios de atención al cliente. Aplicando modelos de análisis avanzados, esta información está sirviendo para decidir lanzar campañas de precios o programas de empleados que consigan aumentar las ventas y aumentar la cuenta de resultados.
- **Desarrollo de las habilidades analíticas:** según se refleja en el estudio, uno de los inhibidores del éxito de Big Data es la falta de formación en la tecnología analítica. Para llegar a tener éxito es importante que las organizaciones inviertan no solo en las herramientas sino también en el modo de utilizarlas para extraer toda la potencia de Big Data.
- **Tener objetivos claros de negocio:** otro desafío subrayado en el estudio es la necesidad de las organizaciones de desarrollar un caso de negocio cuantificable. Hay que destacar los beneficios económicos que se pueden obtener de la retención de clientes, la efectividad de las campañas de marketing, la optimización de la cadena de suministro, el recálculo de precios en tiempo real y del incremento de la productividad de los empleados, al mismo tiempo que se reducen los costes asociados a mantener actualizados los entornos de gestión de la información.

Como lección aprendida, las organizaciones líderes en su sector apuntan que para tener éxito en Big Data es imprescindible que los departamentos de negocio y de tecnologías de la información colaboren y trabajen uno al lado del otro. La aportación no es Big Data, sino la eficiencia, los beneficios e incluso la transformación de negocio que se derivan de su explotación mediante diferentes métodos analíticos.

Para competir en una economía integrada a escala global es evidente que las empresas de hoy en día necesitan una comprensión exhaustiva de los mercados, los clientes, los productos, las normativas, los competidores, los proveedores, los empleados y mucho más. Esta comprensión exige un uso eficaz de información y analítica. De hecho, junto con sus empleados, muchas empresas consideran que la información es su activo más valioso y diferenciador. Ahora, con el surgimiento y la creciente adopción de Big Data, las empresas de todo el mundo están descubriendo formas completamente nuevas de competir y ganar. Se encuentran en proceso de transformación para aprovechar el amplio abanico de información disponible a fin de mejorar la toma de decisiones y el rendimiento en toda la empresa.

Un grupo relativamente pequeño de empresas pioneras ya lo está logrando al proporcionar a sus empleados, desde los altos cargos, pasando por el marketing hasta los trabajadores de menor rango, la información, las habilidades y las herramientas necesarias para tomar decisiones mejores y más oportunas. No todas las organizaciones necesitarán gestionar todo el espectro de funcionalidades. Sin embargo, en todos los sectores existe en cierta medida la posibilidad de utilizar nuevos datos, tecnologías y analíticas. Por ello, independientemente de cuál sea su punto de partida, las empresas de todo el mundo continuarán ampliando el uso de Big Data para obtener valor de negocio y ventajas competitivas en la actual economía globalmente integrada.

# PRÓLOGO

**Jon Ander Gómez Adrián**

**DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN. UNIVERSIDAD  
POLITÉCNICA DE VALENCIA.**



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

El término anglosajón Big Data nos permite dar nombre a la revolución tecnológica que actualmente vivimos respecto a la gestión de los datos. Revolución que es debida, en su mayor parte, al crecimiento exponencial de los datos a medida que avanza el tiempo. Sin embargo, el disponer cada vez de mayores volúmenes de datos nos ha hecho ver que tener datos ya no es sinónimo de poseer información. Máxime cuando éstos no están siempre estructurados ni mucho menos expresados siguiendo reglas fácilmente deducibles. Pensemos en las redes sociales, los datos allí generados suelen estar expresados en lenguaje natural y sin respetar ninguna sintaxis. Este sería un ejemplo de que la extracción de información es una tarea ardua.

Siempre resulta atrevido dar una definición de lo que es Big Data cuando se está inmerso en un periodo de continuos cambios y avances. Desde un punto de vista no tecnológico, Big Data puede definirse como fenómeno socio-económico, pues altera totalmente el modo en que las empresas afrontan la gestión de los grandes volúmenes de datos, y porque hoy en día la mayoría de los datos se generan gracias a la interacción social de un número cada vez mayor de usuarios. Las soluciones Big Data abren nuevas posibilidades, por ejemplo, cuando las empresas pueden combinar su información interna con el conocimiento extraído a partir de los datos de las redes sociales (opiniones sobre productos, tendencias, etc.) no cabe duda que mejoran su toma de decisiones estratégicas.

Desde un punto de vista tecnológico Big Data puede definirse como un conjunto de retos tecnológicos. Retos que aparecen a todos los niveles, desde la infraestructura hardware necesaria para facilitar el almacenamiento y procesamiento de datos hasta el nivel de visualización y representación, pero que tienen su mayor impacto a un nivel intermedio, en los algoritmos y técnicas de procesamiento de grandes volúmenes de datos. Prueba de todo ello son las nuevas herramientas software que han surgido y están surgiendo para explotar de manera extensiva hardware de bajo costo. No cabe duda que todos aquellos procesos que puedan realizarse de manera paralela o distribuida verán reducido sustancialmente el tiempo necesario para completarse.

Partiendo de la definición de Big Data como conjunto de retos tecnológicos, y atendiendo a la amplitud del abanico de herramientas software de reciente aparición y su constante evolución, ya no tiene mucho sentido hablar de un experto Big Data, sino más bien de un equipo multidisciplinar Big Data. Cada miembro de un equipo Big Data responderá a un perfil

de entre los siguientes: administrador de infraestructuras Big Data, responsable de extraer información e intérprete de la información. Los dos últimos perfiles son especializaciones del perfil científico de datos.

El administrador de infraestructuras Big Data debe dominar las herramientas que permiten explotar hardware asequible de manera extensiva, ya sea en las instalaciones de la propia empresa o institución como contratando espacio de almacenamiento y nodos de cómputo en la nube. El responsable de extraer información debe conocer técnicas estadísticas y de aprendizaje automático (*Machine Learning*) para depurar los datos, procesarlos, transformarlos, etc. Todo ello con el fin de extraer la información relevante y obviar la no necesaria. Por último, el intérprete de la información será también experto en utilizar herramientas estadísticas y a la vez que dominará técnicas avanzadas de representación de los datos con el objetivo de transmitir la información extraída de la manera más eficaz posible a quien toma las decisiones en la empresa (consejo de administración, gerente, etc.). El equipo Big Data estará al cargo del proceso de recogida, combinación, procesamiento y gestión de los datos para la toma de decisiones. En otras palabras, el equipo Big Data llevará a cabo todos los pasos necesarios para desarrollar la inteligencia de negocio (*Business Intelligence*) de la organización.

Cada perfil requiere de una formación específica que a su vez necesita de unos conocimientos sólidos en informática y estadística. Por tanto, la especialización en Big Data exige una primera formación universitaria en Informática, Matemáticas u otra titulación técnica relacionada (Ingeniería Industrial o de Telecomunicaciones) para después ampliar estudios mediante un máster en Big Data. Como uno no sabe en qué perfil de los tres desarrollará posteriormente su carrera profesional, la formación de postgrado en Big Data debe cubrir unos mínimos en todas las materias asociadas a cada perfil. Dicho de otro modo, debe facilitar al recién titulado enfocar su carrera hacia el mundo Big Data, y al mismo tiempo debe proporcionar al profesional de las TIC los nuevos conocimientos que le sean necesarios para adaptarse a los avances tecnológicos.

Los nuevos retos tecnológicos que acompañan al Big Data hacen cada vez más necesaria una oferta formativa de postgrado que permita a los profesionales reciclarse y especializarse de manera continua.

## RESUMEN EJECUTIVO

**Big Data** se ha convertido en la **gran tendencia tecnológica y empresarial de los últimos años**. Grandes cantidades de datos (**Volumen**), de múltiples formatos y fuentes (**Variedad**), en continuo movimiento (**Velocidad**) y constante cambio (**Variabilidad**), donde la fiabilidad (**Veracidad**) y la capacidad de representación de los mismos es fundamental (**Visualización**). Sin olvidar los importantes beneficios que posibilita (**Valor**), tanto para las empresas como para los ciudadanos, potenciando el **crecimiento económico** (206.000 millones de euros para la economía europea en 2020), **del empleo** (3,75 millones de empleos en la Unión Europa hasta 2017) **y mejoras sociales** en todos los ámbitos de actividad (transporte, turismo, salud,...).

Este enorme volumen de datos tiene su origen en la gran cantidad de información que está generando el ser humano, sobre todo a raíz de la explosión de los dispositivos móviles conectados a Internet (**más de 1.570 terabytes por minuto**, a través de redes sociales, comercio electrónico, etc.), así como los generados por las máquinas (30 millones de sensores interconectados formando el **Internet de las Cosas**).

Adoptar soluciones empresariales de Big Data está aportando **ventajas competitivas** nada despreciables a dos niveles: la **mejora de la experiencia del cliente** y la **mejora de la eficiencia de los procesos de negocio**, gracias a la toma de **decisiones más inteligentes, más rápidas y que marcan la diferencia**. Por su parte, los ciudadanos se beneficiarán de **servicios mejores y más económicos**, como por ejemplo, un sistema de salud más personalizado, novedosos sistemas de soporte a la decisión para su vida diaria o nuevas formas de interactuar con empresas y Administración Pública.

El gran impacto del Big Data en la **eficiencia, la eficacia, la productividad, la calidad y el valor añadido en los productos y servicios**, entre otros, no está pasando desapercibido. Casi tres cuartas partes de las organizaciones mundiales en 2014 ya estaban invirtiendo en Big Data o lo tienen planificado para los próximos años, destacando la región de Norteamérica y el sector Medios y Comunicación como principal inversor.

Este cambio de paradigma que está propiciando el Big Data va ligado en la mayoría de los casos al **Open Data**, pues la apertura de la información aporta gran riqueza a este análisis inteligente de grandes cantidades de datos, beneficiando tanto a los que poseen los datos como a los que los usan y, si además es realizada por los gobiernos, supone aumentar su transparencia y credibilidad.

La **cadena de valor** de Big Data proporciona un marco que examina cómo reunir datos dispersos de forma organizada y crear información de valor que pueda servir para la toma de decisiones a nivel empresarial, incluyendo así cuatro grandes fases, desde la **generación y adquisición de los datos**, su **procesado y análisis**, la posterior **curación y almacenamiento** de los mismos, para finalmente llegar a su **visualización y a la creación de servicios** asociados. A lo largo de esta cadena de valor, las empresas/entidades usuarias, los generadores y proveedores de datos, los proveedores tecnológicos y los proveedores de servicios se configuran como los actores clave, sin olvidar el importante papel que juegan las entidades reguladoras, las de estandarización y las redes colaborativas/ecosistemas.

En el plano científico-tecnológico, las **áreas técnicas prioritarias** donde se debe desarrollar I+D en el entorno del Big Data son: **Ingeniería y gestión de datos, Arquitecturas optimizadas para analítica de datos estáticos y dinámicos, Analítica avanzada con algoritmos predictivos y prescriptivos, Visualización y experiencia de usuario avanzada, y Mecanismos de privacidad y anonimización.** En este sentido, ya existe actualmente todo un **ecosistema de tecnologías** de Big Data agrupados en cuatro grandes categorías: *tecnologías base, infraestructura, análisis y aplicaciones*, que permitirá solucionar los distintos problemas o retos que se tengan mediante la orquestación de varias de ellas.

En cuanto al dominio de aplicación, son numerosos los **sectores empresariales** que ya se están beneficiando del Big Data o podrían hacerlo: *salud*, posibilitando una **medicina personalizada, predictiva, preventiva y participativa**, donde la aplicación de sistemas cognitivos supone un cambio radical en los procesos de diagnóstico; *turismo*, focalizándose en la **personalización de los servicios turísticos**, creando experiencias enriquecedoras y adaptadas así como generando **nuevos modelos de negocio**; *fabricación*, siendo uno de los pilares de las **fábricas 4.0 que garantizan sostenibilidad, excelencia operativa y eficiencia**; *banca y finanzas*, donde el Big Data podría **mejorar la visión integral de los clientes y la gestión de riesgos y de fraudes**, propiciando mayor eficiencia en la provisión de servicios financieros y mayor satisfacción de los clientes; *gestión urbana y ciudades inteligentes*, donde el Big Data aportaría mejoras en cuanto a **seguridad ciudadana, transporte urbano, gestión del agua, energía, residuos urbanos, educación y buen gobierno**, creando **ciudades más habitables y sostenibles**; y *medios de comunicación y entretenimiento*, donde se posibilitaría ofrecer **contenidos más personalizados**, mejorando además la relación con los anunciantes, caminando hacia lo que se conoce como **periodismo de datos**.

Existen **diversos obstáculos** a los que hacer frente para una adopción extendida del Big Data, incluyendo la adquisición de **habilidades técnicas, analíticas, de liderazgo y organización**, el disponer de **planes de inversión y financiación**, así como el encontrar **casos de negocio convincentes**, sin olvidarnos de la **disponibilidad de los datos**, los aspectos legales en su uso (**privacidad y seguridad**) así como la **concienciación** sobre los beneficios y el valor que el Big Data puede crear en el sector empresarial, el sector público y para los ciudadanos.

En este escenario son numerosas las entidades, asociaciones, plataformas tecnológicas, iniciativas y proyectos de I+D+i que se están llevando a cabo, donde el Instituto Tecnológico de Informática, miembro fundador de la *Big Data Value Association*, lleva varios años posicionándose y tiene muchas capacidades científico-tecnológicas desarrolladas en tres ámbitos principalmente: plataformas e infraestructuras que soportan el manejo de grandes cantidades de datos, análisis de información (*Data Science*) y visualización orientada a Estrategia/Valor (*Big Data Visualization & Insights*).

# INTRODUCCIÓN

## 1.1 Justificación y objetivos del estudio

El presente documento es un informe de **análisis del estado actual y de las tendencias del Big Data**, como nuevo activo en la economía europea, que propiciará la creación de oportunidades de negocio para las empresas en numerosos ámbitos de aplicación (salud, turismo, fabricación, etc.), a la vez que se maximizan los beneficios para los ciudadanos.

Así, el informe parte de la descripción del **contexto del Big Data**, revisando el significado del término, el origen de todos esos datos, las ventajas que supone adoptar tecnología y aplicaciones Big Data, el grado de adopción empresarial actual así como las primeras cifras de impacto económico general del Big Data, en términos de crecimiento económico y de empleo. Termina esta sección con una introducción al Open Data, tendencia paralela y relacionada con el Big Data.

En un segundo apartado se revisa la **cadena de valor del Big Data** y los principales actores en la misma, para posteriormente detallar las áreas científico-tecnológicas prioritarias para el desarrollo del Big Data así como **el ecosistema de tecnologías existentes** actualmente estructuradas en cuatro grandes categorías: tecnologías base, infraestructuras, análisis y aplicaciones.

En la parte central del documento se recopilan los **agentes e iniciativas relevantes** en este ámbito, principalmente proveedores, plataformas tecnológicas y proyectos europeos de I+D+i recientes. A continuación se revisan los **principales sectores de aplicación** del Big Data así como los **impactos sectoriales cualitativos y cuantitativos** del mismo, incluyendo las principales **barreras de entrada** en su implantación.

Se cierra el informe con las **capacidades y experiencia del Instituto Tecnológico de Informática (ITI)** en el ámbito del Big Data. Este informe ha sido llevado a cabo por el Instituto Tecnológico de Informática dentro de la línea de **Vigilancia Tecnológica** del Departamento de Inteligencia Estratégica y Competitiva. Este departamento de ITI tiene por misión **captar y analizar las tendencias y evolución de las TIC, y las estrategias de investigación y políticas nacionales e internacionales en torno a ellas**, para orientar al conjunto del Instituto y a las empresas sobre nuevos proyectos, productos o procesos que puedan constituirse en oportunidades de I+D+I o negocio, y coordinar la respuesta ante estas oportunidades.

Para el cumplimiento de su misión, el departamento se divide en varias líneas de actividad, configurándose la línea de Vigilancia Tecnológica como una de las más importantes, dado el elevado dinamismo del sector en el que el Instituto se ubica y la importancia de tener una buena información objetiva de partida para poder tomar decisiones estratégicas.

Dentro de la línea de Vigilancia Tecnológica, el Instituto puso en marcha en 2008, gracias al apoyo de IMPIVA y del Fondo Europeo de Desarrollo Regional, el **Observatorio Tecnológico del Sector TIC (<http://observatorio.iti.upv.es>)**, con el objetivo de recopilar, filtrar, categorizar, analizar y difundir información valiosa, y así **impulsar la actividad innovadora y crear inteligencia competitiva** en el Instituto y en las empresas del sector, mejorando por tanto su posición competitiva.

El presente informe se enmarca dentro de las actividades de análisis de tendencias y pretende dar una visión de las tecnologías, proyectos e iniciativas que existen actualmente, además de mostrar la evolución acaecida durante los últimos años, de forma que las empresas cuyo objetivo sea prestar servicios o desarrollar productos en ese ámbito puedan posicionarse, conozcan a los actores más relevantes en el mismo y puedan establecer sus estrategias de futuro.

## 1.2 Contexto del Big Data

### 1.2.1 ¿Qué es Big Data?

Según IBM, uno de los principales *players* en este campo, Big Data es un término que describe **enormes cantidades de datos que no pueden ser procesados o analizados usando procesos o herramientas tradicionales**. En cuanto a la cantidad, en general empieza a utilizarse el término Big Data cuando se habla de *petabytes* de datos, aunque lo que constituye un volumen “alto” varía en función del sector e incluso de la ubicación geográfica<sup>1</sup>.

Además del volumen, que es la característica que se asocia con mayor frecuencia a Big Data, existen otras dimensiones que convergen en su caracterización, es lo que se conoce como **las 4 V del Big Data: Volumen, Variedad, Velocidad y Veracidad**, tal y como muestra la figura siguiente del informe “Analytics: el uso de big data en el mundo real” de IBM [1].

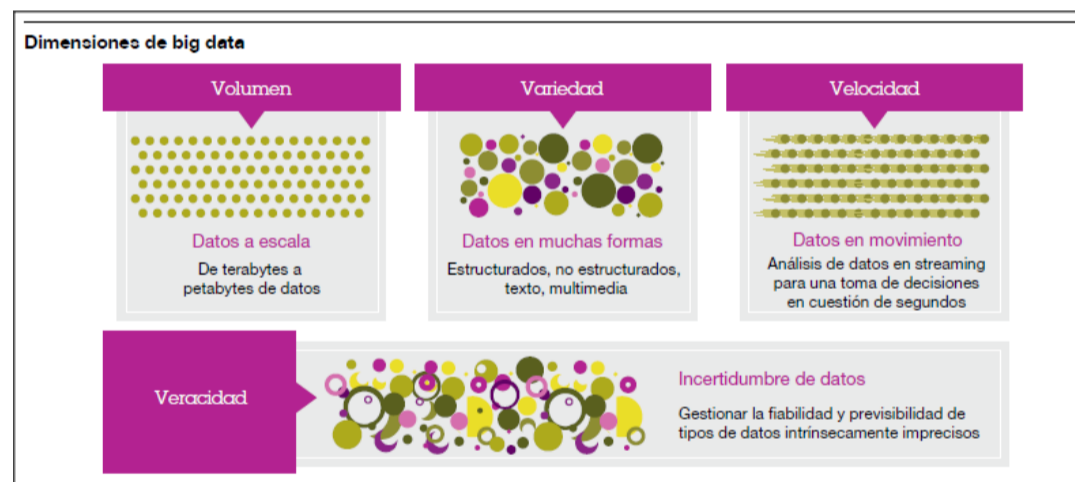


Figura 1. Dimensiones de Big Data (Fuente: IBM [1])

La variedad hace referencia a diferentes tipos y fuentes de datos, incluidos los **datos estructurados, semi-estructurados y no estructurados**, teniendo en cuenta que los datos que se generan actualmente presentan muchas formas: texto, audio, vídeo, datos web, datos procedentes de redes sociales, datos de sensores, archivos de registro, etc.

Por otro lado, cada vez es mayor la velocidad a la que se crean, procesan y analizan los datos. Cuanto mayor velocidad, menor tiempo de espera entre el momento en el que se crean los datos, el momento en el que se captan y el momento en el que están accesibles. Hoy en día los datos se generan a una velocidad que los sistemas tradicionales no pueden manejar. Además, en algunos procesos el tiempo es fundamental, se necesita **análisis en tiempo real**, es decir, que la velocidad de respuesta sea lo suficientemente rápida como para lograr obtener la información correcta en el momento preciso.

1 <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>. Gigabyte =  $10^9$ ; Terabyte =  $10^{12}$ ; Petabyte =  $10^{15}$ , Exabyte =  $10^{18}$ .

La veracidad hace referencia al **nivel de fiabilidad asociado a ciertos tipos de datos**. Un reto y requisito fundamental del Big Data es que los datos tengan alta calidad, sin embargo, hay ciertos datos que son intrínsecamente inciertos, por ejemplo, los sentimientos y la sinceridad de los seres humanos, los sensores GPS que rebotan entre los rascacielos, las condiciones climáticas, los factores económicos y el futuro. A pesar de la incertidumbre, estos datos contienen información valiosa.

Dando un paso más allá, destacamos el planteamiento que está realizando la **Big Data Value Association (BDVA)**<sup>2</sup>, organización sin ánimo de lucro auto-financiada formada por 24 entidades de referencia, de la industria de los macrodatos como IBM, ATOS, INDRA, Nokia, Orange, SAP o Siemens, y de la investigación en TIC, en donde se encuentra el **Instituto Tecnológico de Informática**<sup>3</sup>, el de Fraunhofer y universidades como la Bolonia (CINI) y la Universidad Politécnica de Madrid. El objetivo de este organismo es establecer la **agenda estratégica de investigación en el campo del Big Data**, para crear un ecosistema de innovación que permita a Europa alcanzar y mantener su liderazgo en dicho ámbito, a la vez que se maximizan los beneficios sociales y económicos europeos (de sus negocios y sus ciudadanos).

En concreto, la BDVA añade el Valor como la **5ª V de caracterización del Big Data**: Volumen, Velocidad, Variedad, Veracidad y **Valor**. Dominando la generación de Valor de Big Data se creará una ventaja competitiva para la industria europea, proporcionando crecimiento económico y de empleo.

Por su parte, el *Online Business School* (OBS) en su reciente estudio "Big Data 2015" [2], resalta que actualmente no sólo hay que tener en cuenta los retos ya comentados de Volumen, Variedad, Velocidad y Veracidad, sino que también es necesario atender a la **Variabilidad, la Visualización y el Valor**, dando a paso a las **nuevas 7 Vs** como principales **principios directores de cualquier proyecto Big Data**.

Según el citado estudio, la variabilidad se refiere a los datos cuyo significado está en constante cambio, situación que ocurre especialmente cuando la recolección de datos se basa en el procesamiento del lenguaje. La visualización hace referencia a la manera de representar la información. Con la gran cantidad de datos que son necesarios ahora, es necesario disponer de formas más accesibles y fáciles de leer esos datos. Por último, el valor potencial del Big Data es muy grande, pero el coste almacenar mala o información pobre, también es muy grande. Por ello se debe tener presente que el valor reside en un análisis riguroso de datos precisos, y en la información y los conocimientos que de aquí se pueden sacar.

En definitiva, Big Data es una combinación de estas características que crea una oportunidad para que las empresas puedan obtener una **ventaja competitiva**. Big Data permite a las empresas transformar la forma en la que interactúan con sus clientes y les prestan servicio, y posibilita la transformación de las mismas e incluso de sectores enteros.

## 1.2.2 ¿De dónde provienen todos estos datos?

Los seres humanos estamos creando y almacenando información continuamente y cada vez más en cantidades enormes. Según el OBS, **en los últimos 10 años se ha creado más**

2 Big Data Value Association <http://www.bdva.eu/>

3 <http://www.iti.es/es/actualidad/noticias/17119/el-instituto-tecnologico-de-informatica-la-vanguardia-del-big-data-europeo>

**información que en toda la historia de la humanidad**, hecho motivado principalmente por el desarrollo de los **dispositivos móviles con conexión a Internet**, del **comercio electrónico** y de las **redes sociales**<sup>4</sup>.

Destacan que el volumen de datos generados en 2014 se ha multiplicado. En un minuto, en Internet se generan 4,1 millones de búsquedas en Google, se escriben 347.000 tuits, se comparten 3,3 millones de actualizaciones en Facebook, se suben 38.000 fotos a Instagram, se visualizan 10 millones de anuncios, se suben más de 100 horas de vídeo a YouTube, se escuchan 32.000 horas de música en *streaming*, se envían 34,7 millones de mensajes instantáneos por Internet o se descargan 194.000 apps. En total, **más de 1.570 terabytes de información por minuto**<sup>5</sup>.

En cuanto a los dispositivos móviles, según OBS, estos serán el mayor proveedor para los sistemas de Big Data. Se estima que, en 2020, **más de 30.000 millones de dispositivos estarán conectados a Internet**, un incremento exponencial comparado con los cerca de 2,5 billones que había en 2009.

Pero no solamente los seres humanos contribuyen a este crecimiento enorme de datos, existe también la comunicación denominada máquina a máquina (**M2M, machine-to-machine**) cuyo valor en la creación de grandes cantidades de datos también es muy importante. Según IBM, hay más de 30 millones de sensores interconectados en distintos sectores como automoción, transporte, industria, servicios, etc. y se espera que este número crezca en un 30% anualmente. Este ecosistema forma lo que se conoce como el **Internet de las Cosas (IoT-Internet of Things)**.

En la siguiente figura se resumen los distintos tipos de datos de Big Data<sup>6</sup>:

- *Web and social media*: contenido web e información obtenida a través de redes sociales como Facebook, Twitter, LinkedIn, etc. y blogs.
- *Machine-to-Machine*: M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas, etc.), los cuales transmiten a través de redes (inalámbricas, híbridas, cableadas...) a otras aplicaciones que traducen estos eventos en información significativa.
- *Big Transaction Data*: registros de facturación, de llamadas telefónicas, etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.
- *Biometrics*: información biométrica como huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. Estos datos biométricos son importantes en áreas de seguridad e inteligencia.
- *Human Generated*.

4 <http://www.obs-edu.com/noticias/informe/el-volumen-de-datos-generado-por-smartphones-crecera-un-63-los-proximos-cuatro-anos/>

5 <http://www.obs-edu.com/noticias/estudio-obs/en-2020-mas-de-30-mil-millones-de-dispositivos-estaran-conectados-internet/>

6 <http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/>

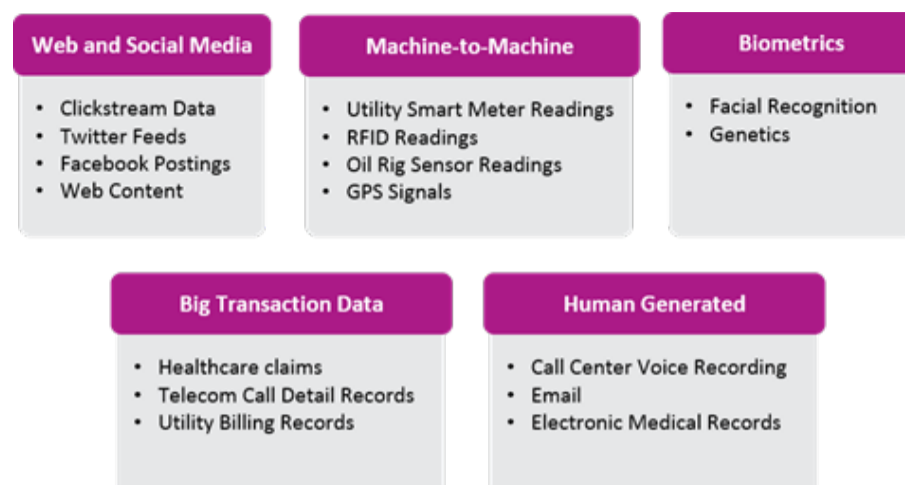


Figura 2. Tipos de datos de Big Data (Fuente: Sunil Soares, Information Asset, LLC, 2012)

### 1.2.3 ¿Qué ventajas supone adoptar Big Data?

Según el 63% de los profesionales entrevistados por IBM de 95 países y 26 sectores en el citado estudio "Analytics: el uso de big data en el mundo real", el uso de la información y la analítica está dando lugar a una ventaja competitiva para las empresas.

Consideran que Big Data proporciona la capacidad para **comprender y predecir mejor los comportamientos de los clientes** y, al hacerlo, **mejorar su experiencia**. A través de esta comprensión profunda, empresas de todo tipo encuentran **nuevas formas de interactuar con sus clientes actuales y futuros**. Este principio es aplicable al **comercio minorista**, pero también a **las telecomunicaciones, la sanidad, el gobierno, la banca y las finanzas, y al sector de productos al consumidor**, donde usuarios finales y ciudadanos están involucrados en interacciones *business-to-business* (B2B) entre socios y proveedores.

Con el surgimiento y la creciente adopción de Big Data las empresas están descubriendo formas completamente nuevas de competir y ganar. Se encuentran en proceso de transformación para aprovechar el amplio abanico de información disponible a fin de **mejorar la toma de decisiones y el rendimiento** en toda la empresa. Es decir, mediante la adopción de estrategias de Big Data las empresas serán capaces de prestar un mejor servicio a sus clientes y de mejorar sus operaciones, **aumentando su productividad y su competitividad**, en base a:

- **Decisiones más inteligentes** – Aprovechar nuevas fuentes de datos para mejorar la calidad de la toma de decisiones.
- **Decisiones más rápidas** – Permitir una captura y análisis de datos en tiempo más real para respaldar la toma de decisiones en el "punto de impacto", por ejemplo, cuando un cliente está navegando por un sitio web o al teléfono con un representante del servicio de atención al cliente.
- **Decisiones que marquen la diferencia** – Centrar las iniciativas de Big Data en ámbitos que proporcionen una verdadera diferenciación.

Por su parte, el citado estudio del OBS indica que las áreas donde el Big Data aporta más valor son en la **mejora de la experiencia de cliente** y en la **mejora de la eficiencia de los procesos de negocio**, además de en campos como el **transporte, la salud, los medios de comunicación, la banca, las comunicaciones y el comercio minorista**. Big Data da la oportunidad de que las empresas tengan **mayor capacidad de decisión**, decisión que hasta ahora se apoyaba más en la intuición. Esto es un cambio de paradigma, que implica aprovechar todas las infraestructuras, todas las herramientas y una nueva forma de pensar para poder **introducir líneas de negocio nuevas** hasta ahora impensables.

#### 1.2.4 ¿Qué grado de adopción empresarial tiene el Big Data?

**En 2014, el 73% de las organizaciones mundiales estaban invirtiendo o tenían planificado invertir en Big Data en los próximos años**, según el OBS, lo que supone un incremento del 5% respecto al año anterior. Se indica que el porcentaje de compañías que no tenía previsto incorporar proyectos de Big Data a sus planes estratégicos se ha reducido respecto al año anterior, pasando del 31% al 24%. Destacan también que **la región de Norteamérica es la que está liderando la inversión y adopción** de proyectos y herramientas de Big Data. Así mismo, el informe indica que en 2014 se registra un **notable aumento de las inversiones en todos los sectores empresariales**, siendo Medios y Comunicación el principal inversor, donde el 55% de las organizaciones ya han realizado inversiones y un 33% lo tienen en sus planes.

Por otro lado, en el estudio de IBM los entrevistados sugieren **cuatro fases** principales en el proceso de **adopción y evolución de Big Data**, tal y como muestra la figura posterior:

- La fase de **educación** se centra en la concienciación y el desarrollo del conocimiento, puesto que los entrevistados indicaban que los directivos empresariales aún no habían comprendido totalmente ni abrazado el potencial del Big Data.
- En la fase de **exploración** se trata de desarrollar la hoja de ruta de la empresa para el desarrollo del Big Data, teniendo en cuenta los datos, la tecnología y las habilidades existentes.
- En la fase de **interactuación**, las empresas comienzan a comprobar el valor de negocio de Big Data, así como a llevar a cabo una valoración de sus tecnologías y habilidades.
- En la fase de **ejecución**, el nivel de operatividad e implementación de las funciones analíticas y de Big Data es mayor dentro de la empresa. Sin embargo, sólo el 6% de los encuestados confirmó que sus empresas habían implementado dos o más soluciones de Big Data a escala.

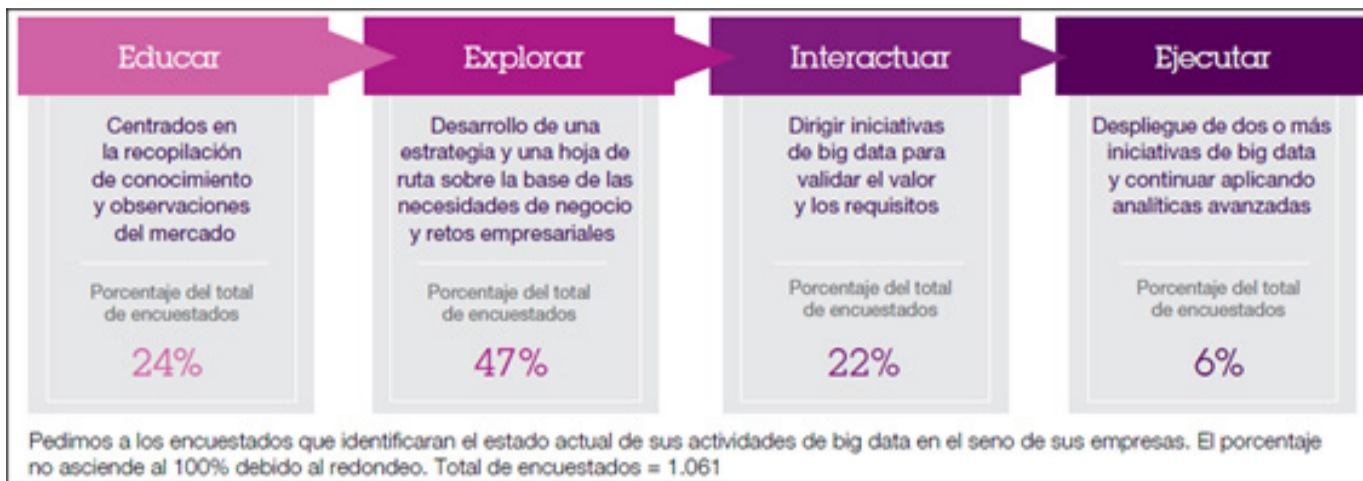


Figura 3. Fases del patrón de la adopción de Big Data (Fuente: IBM [1])

Por último, es interesante también recoger el análisis de la consultora Gartner, que sitúa Big Data en “**Peak of Inflated Expectations**”, tal y como muestra la siguiente figura, es decir, existen casos de éxito acompañados por una serie de fracasos en este tipo de proyectos<sup>7</sup>.



Figura 4. Hiper ciclo para Tecnologías Emergentes 2014 (Fuente: Gartner)

### 1.2.5 ¿Qué impacto supondrá el Big Data?

El desarrollo del Big Data impactará en la **competitividad de las empresas** y, por ende, resultará en un **crecimiento económico y de empleos** a nivel mundial. Por su parte, los ciudadanos se beneficiarán de **servicios mejores y más económicos**, como por ejemplo, un sistema de salud más personalizado, novedosos sistemas de soporte a la decisión para su vida diaria o nuevas formas de interactuar con empresas y Administración Pública, todo ello

7 Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business <http://www.gartner.com/newsroom/id/2819918>

basado en soluciones Big Data.

A nivel cuantitativo y focalizándonos en Europa, según indica el reciente estudio del OBS, la **transición a soluciones tecnológicas basadas en datos supondrá una aportación de 206.000 millones de euros para la economía de Europa en el año 2020**, lo que repercutirá en un **incremento en el PIB de la Eurozona de un 1,9%**, cifra equivalente a un año completo de crecimiento económico en la Unión Europea. En función de las distintas zonas geográficas, se indica que este **crecimiento del PIB es del 2,2% en el Norte de Europa, del 1,9% en nuevos estados miembros y del 1,6% en el Sur de Europa**. Así mismo, los sectores económicos más beneficiados son **Comercio (47.000 millones de euros), Industria (45.000 millones), Administración Pública (27.000 millones) y Sector Sanitario (10.000 millones)**.

En cuanto a la generación de empleo por la adopción de Big Data, según estimaciones realizadas por la BDVA, se espera que se **creen 3,75 millones de empleos en la Unión Europea hasta 2017**.

### 1.2.6 Big Data versus Open Data

En general se observa cómo la **apertura de los datos es complementaria al cambio hacia las soluciones de Big Data**. Ya en 2012 la consultora Gartner en su estudio "Open for Business: Learn to Profit by Open Data" llega a la conclusión de que, si bien el análisis de grandes volúmenes de datos vuelve a las empresas más inteligentes, la riqueza real está en la apertura de información<sup>8</sup>.

Un dato o contenido es abierto si cualquiera puede usarlo, reutilizarlo y redistribuirlo. Esto significa que cualquier parte interesada puede acceder a los datos sin barreras adicionales de coste, burocracia, técnicas, o de uso. No sólo se trata de dar datos de forma gratuita sino que implica los siguientes aspectos, tal y como destaca el informe "Big Data & Open Data in Europe. A growing engine or a missed opportunity", del Warsaw Institute for Economic Studies [3]:

- Cambio de enfoque desde "la apertura por petición" a la "**apertura por defecto**".
- Provisión de datos en **formatos apropiados**, fáciles de analizar y modificar, con las limitaciones en su uso que permita la ley.
- Provisión de los datos junto a su descripción (**metadato**) que ayude al usuario a valorar las fortalezas y debilidades de esos datos así como a entender de forma precisa su significado.
- Provisión de los datos bajo una **licencia abierta**, que no imponga ninguna restricción en el uso.
- Publicación de los **datos primarios** y en su estado original si es posible.
- Mantenimiento del valor de los datos.
- Establecimiento de un **punto de contacto** para los usuarios, que sirva de ayuda y consulta y, a la vez, de recepción de feedback para mejorar el acceso a los mismos.

Aunque el debate actual del Open Data se centra en el sector público, los gobiernos no son las únicas fuentes posibles de datos abiertos, también pueden proceder de empresas, instituciones de investigación y de las propias personas a nivel individual.

Según el citado estudio, abrir y facilitar el acceso a los datos por todas las partes proporciona beneficios tanto para los que tienen los datos como para los que los usan. Si además se lleva a cabo por parte de los gobiernos, supone aumentar su transparencia y credibilidad. Así mismo, cuanto mayor es el grado de apertura de los datos, mayor es el tamaño y la diversidad de la comunidad que accede a ellos. Así, y de acuerdo al efecto multiplicador de compartir en la red, el valor de un *dataset* se incrementará cuanto más fácil sea acceder a él y reutilizarlo. Por contra, existen importantes razones para no abrir los datos como la privacidad de las personas, el mantenimiento de las ventajas competitivas de las empresas así como el aseguramiento de una competencia justa, y finalmente aspectos relativos a la seguridad nacional. En la siguiente figura se resume este escenario de coste-beneficio de la apertura de los datos para sus propietarios.

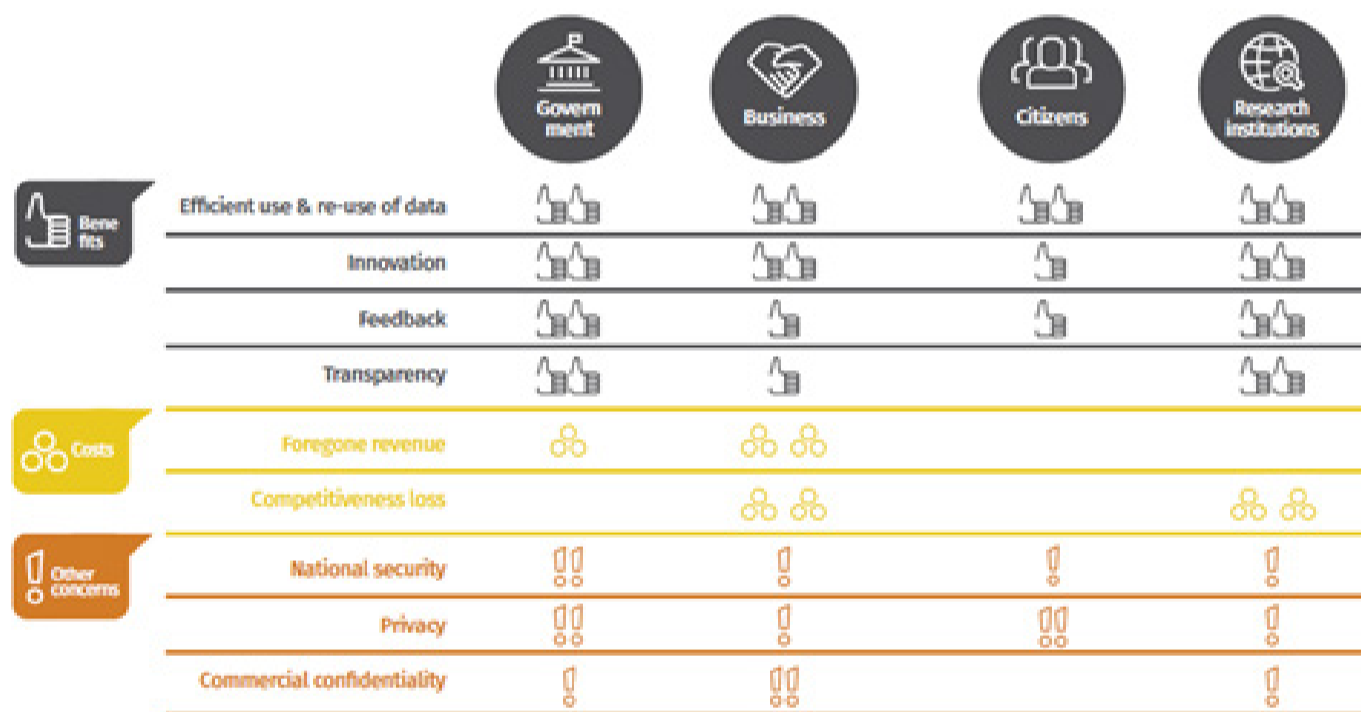


Figura 5. Coste-beneficio de la apertura de los datos para sus propietarios (Fuente: WISE Institute, [3])

## CADENA DE VALOR Y TECNOLOGÍAS IMPLICADAS

### 2.1 Cadena de valor y principales “stakeholders”

Hace más de una década, Michael E. Porter<sup>3</sup> introdujo el concepto de cadena de valor, que consiste en un modelo teórico que permite describir el desarrollo de una serie de actividades que crean y construyen valor, valor que en última instancia una organización empresarial ofrece a sus clientes.

De forma similar a la caracterización de Big Data a través de las 7 Vs descritas en el apartado anterior de introducción, podríamos hablar de las **3 Cs de la cadena de valor de Big Data: Colección, Consolidación y Consumo**, que definen a alto nivel la forma de administrar los datos, desde su generación y adquisición hasta su transformación y representación para su aprovechamiento por los usuarios<sup>4</sup>.

En línea con lo anterior, según el artículo “From Data to Decisions: A Value Chain for Big Data” [4], la cadena de valor de Big Data proporciona un marco que examina **cómo reunir datos dispersos de forma organizada y crear información de valor que pueda servir para la toma de decisiones a nivel empresarial**. En concreto, se definen 3 grandes bloques, tal y como muestra la figura posterior:

- **Descubrimiento de los datos**, que incluye el **inventariado de las fuentes** de datos disponibles, su **preparación** (establecimiento del acceso a las fuentes de datos y de las reglas de control de acceso para garantizar la seguridad y privacidad), y la **organización** (identificación de la sintaxis, estructura y semántica de cada fuente de datos).
- **Integración de los datos**, donde se establece una representación común de los datos, que permitirá su análisis particular posteriormente. La integración puede ser virtual, a través de modelos federados, o física, a través de un repositorio de datos. Las tecnologías tradicionales de federación de datos y las emergentes tecnologías de web semántica soportarán la integración y consulta de fuentes de datos combinadas.
- **Explotación de los datos**, que implica el aprovechamiento de los datos recogidos e integrados en las fases anteriores para la **toma de decisiones informadas**. El objetivo es proporcionar a los agentes relevantes información útil en el formato adecuado para la toma de decisiones críticas. El punto final de la cadena de valor es la determinación de las acciones necesarias teniendo en cuenta los resultados visualizados.

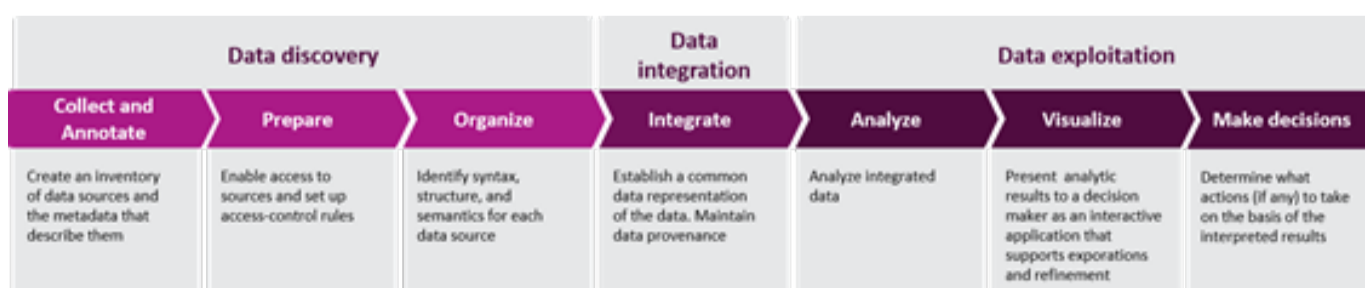


Figura 6. The data value chain (Fuente: IT Professional Magazine, IEEE Computer Society [4])

3 Michael E. Porter <http://www.hbs.edu/faculty/Pages/profile.aspx?facId=6532>

4 Big Data: Evaluando impacto y aceleración en las industrias <http://download.sap.com/latinamerica/download.epd?context=E-9919DC613E7CB242D5B2D72F4610A8272EA8387BB90F05F1CA8F70C8797FEF74A46C8996AB86E75E8178C02DBD108FC7BF7B7BEFE669903>

Otra clasificación que se encuentra disponible en la bibliografía considera cuatro fases fundamentales para la cadena de valor de Big Data y lo ejemplariza de forma sencilla a través del caso de uso de datos de tráfico GPS<sup>5</sup>, tal y como muestra la figura siguiente:

- **Generación de datos**, que incluye la recolección de los datos desde la fuente que los origina. En el caso de los datos de tráfico, tradicionalmente estos eran generados por los sensores instalados en la carretera, que son propiedad de los gobiernos locales y municipales. Posteriormente, a través de compañías de desarrollo de aplicaciones se ha posibilitado la existencia de plataformas de recolección colaborativa de datos de tráfico (como la aplicación Waze).
- **Almacenamiento de datos**, donde los datos en bruto son almacenados y combinados con datos de otras fuentes para su preparación para la posterior interpretación. En el caso de los datos de tráfico, las autoridades locales típicamente almacenan los datos en sus propios servidores o lo externalizan a una tercero (compañía de repositorio de datos).
- **Motor de inteligencia**, fase en la que se aplica algoritmia y analítica para obtener valor de los datos. En el caso de uso ejemplo, Garmin y otros fabricantes de GPS confían en los motores de inteligencia avanzados de NAVTEQ (y sus competidores) para interpretar los datos de sensores de tráfico.
- **Uso de los datos**, etapa final de la cadena de valor en la que la salida del motor de inteligencia se convierte en valor tangible. Este valor puede proceder de la experiencia directa del usuario, como es el caso de una persona que usa Garmin y recibe la notificación de que su viaje se retrasará 10 minutos, o de la comunicación con sistemas inteligentes, donde las señales de tráfico se ajustan automáticamente basándose en el flujo de tráfico real.

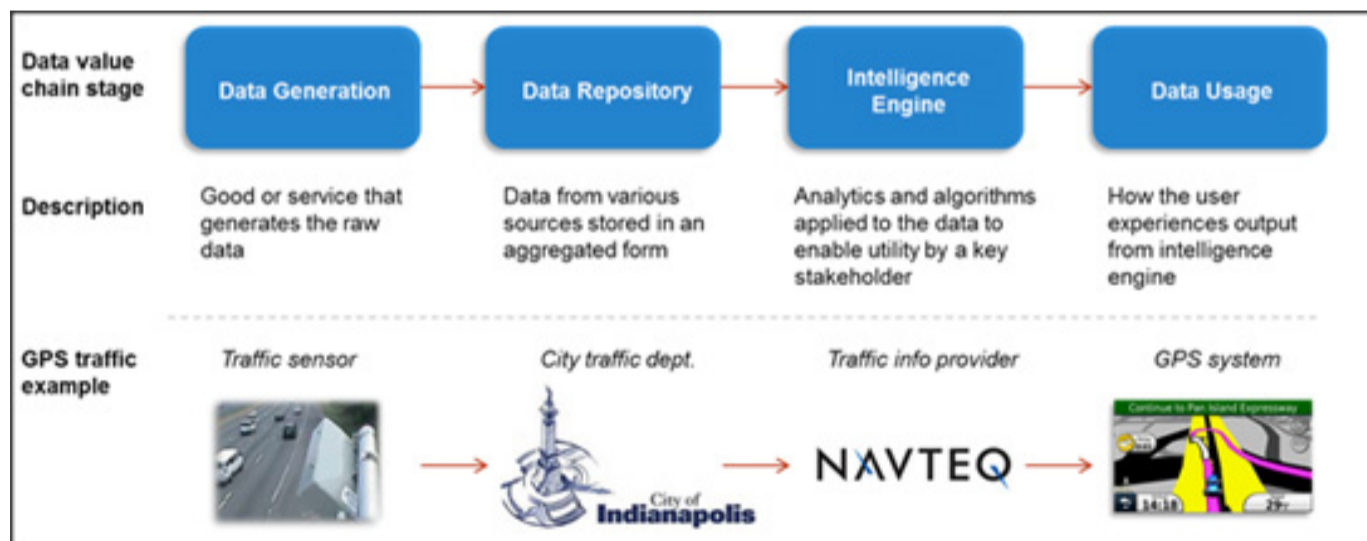


Figura 7. Cadena de valor Big Data para el caso de uso de datos de tráfico (Fuente: Innosight, consultora estratégica y de innovación, 2013)

Destacamos también otro ejemplo interesante de la cadena de valor para obtener conocimiento de los datos. Es el que realiza HP en el contexto de los proveedores de servicios de comunicaciones (CSPs) en su artículo "From big data to knowledge: value chain for CSPs" [5].

En este caso de uso, Big Data permitiría a los CSPs crear inteligencia para operar más eficientemente la red, analizar el éxito de los servicios que están ofreciendo y mejorar la

5 Winning within the Data Value Chain <http://www.innosight.com/innovation-resources/strategy-innovation/winning-within-the-data-value-chain.cfm>

experiencia personal de sus clientes. Los CSPs conocen una ingente cantidad de información personal de sus clientes, como los números móviles de sus contactos, las direcciones físicas, el uso de Internet, la descarga de aplicaciones, el historial de navegación, e incluso hasta cuánto dura su trayecto al trabajo cada mañana. Así mismo, la red de comunicaciones, los sistemas de tarificación, y los distintos departamentos de servicio han ido generando durante años cientos de millones de registros de los detalles de las llamadas individuales. Según HP, para que la toma de decisiones de negocio sea mejor, los CSPs deben considerar la cadena de valor completa, que transforme sus datos en conocimiento, lo que incluye:

- **Fuentes de datos**, lo que incluye la identificación de las distintas fuentes de Big Data, como por ejemplo, información de la red, de los sistemas de facturación, de los perfiles de los suscriptores, de sus dispositivos o de redes sociales.
- **Colecciones de datos**, incluyendo diferentes tecnologías como las sondas de red que capturan los datos.
- **Gestión y estructuración de los datos**, donde se debe proveer de rápido acceso a los datos.
- **Acceso a los datos**, que permite búsquedas interactivas.
- **Business Intelligence**, que consiste en el proceso de analítica aplicado a un número de casos de uso específicos de los CSPs.
- **Presentación y visualización**, donde se proponen predicciones y resultados a la plantilla en un formato usable.

Por último, la Big Data Value Association, de la que ITI es socio fundador, en su propuesta para la creación de una colaboración público-privada contractual (**cPPP - contractual Public Private Partnership**) en Big Data, ha desarrollado una cadena de valor en este ámbito que sirve de marco para esbozar las tecnologías implicadas [6], tal y como muestra la figura siguiente:

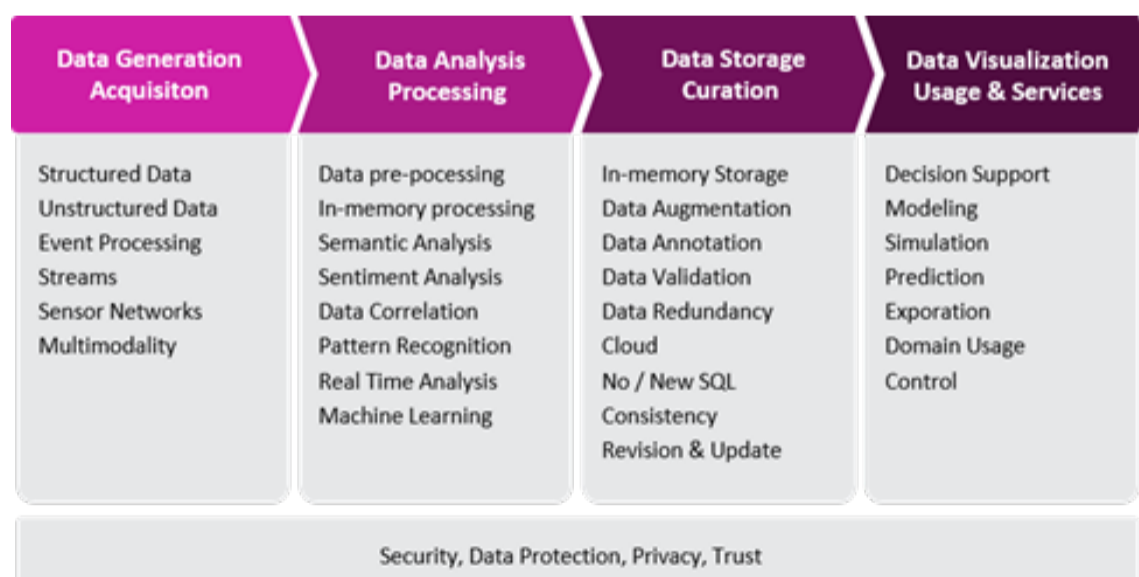


Figura 8. Ecosistema de Big Data a lo largo de la cadena de valor (Fuente: BDVA [6])

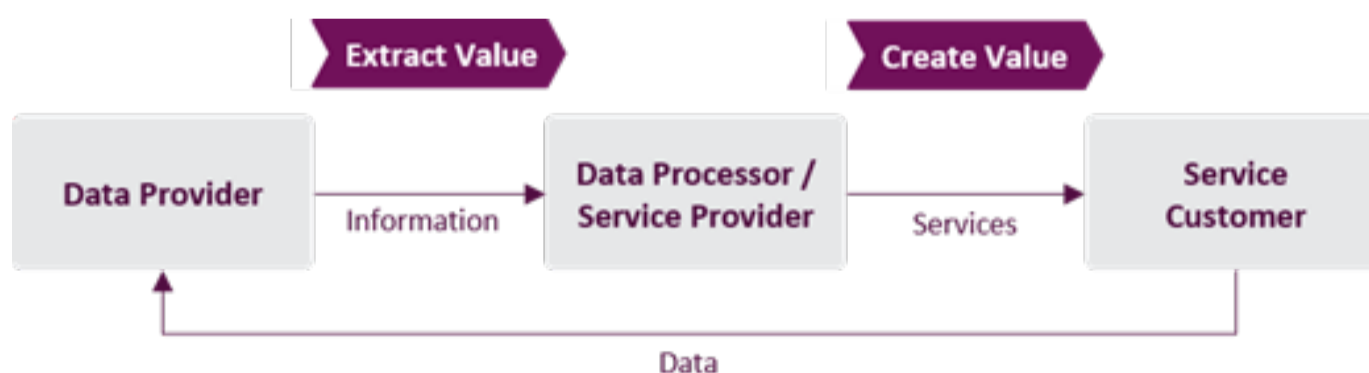
Se observa cómo la cadena de valor propuesta, al igual que las mencionadas anteriormente, se basa en un conjunto de fases que van desde la **generación y adquisición de los datos**, su **procesado y análisis**, la posterior **curación y almacenamiento** de los mismos, para finalmente llegar a su **visualización y a la creación de servicios** asociados.

Según este análisis de la BDVA, ya existen bastantes empresas en Europa que proporcionan servicios y soluciones dentro de este ecosistema de Big Data. Algunas generan y dan acceso a grandes cantidades de datos, tanto estructurados como no estructurados. Típicamente adquieren o combinan flujos de datos en tiempo real de distintas fuentes y añaden valor mediante pre-procesado, validación y aseguramiento de la integridad. Otras se especializan en analizar los datos e identificar correlaciones y patrones. Así mismo, hay empresas que usan este conocimiento para hacer predicciones y para la toma de decisiones en distintos dominios de aplicación.

Por otro lado, este ecosistema se compondrá de nuevos agentes, derivados de los nuevos modelos de negocio que se crearán alrededor de los nuevos conceptos o técnicas de recolección, procesado, análisis, visualización y uso de los datos que se generarán en los próximos años.

En definitiva, los **actores clave** a lo largo de esta cadena de valor de Big Data según la BDVA son:

- **Empresas/entidades usuarias**, organizaciones en todos los dominios, condiciones (públicas, privadas, con o sin ánimo de lucro) y de todos los tamaños que quieran mejorar sus procesos, sus servicios o productos usando tecnología Big Data y productos y servicios basados en datos.
- **Generadores y proveedores de datos**, que crean, recolectan, agregan, transforman y modelan los datos en bruto de varias fuentes, públicas y privadas, y los ofrecen a sus clientes.
- **Proveedores tecnológicos**, que proporcionan herramientas y plataformas para la gestión y analítica de datos, extrayendo conocimiento de los mismos, "curándolos" y visualizándolos.
- **Proveedores de servicio**, que desarrollan aplicaciones Big Data sobre las herramientas y plataformas para proporcionar servicios a las empresas usuario.



*Figura 9. Actores principales del ecosistema de Big Data (Fuente: BDVA [6])*

Además de lo anterior, existen un conjunto de organizaciones y comunidades que tendrán impacto en este ecosistema y que tienen influencia sobre toda la cadena de valor:

- **Entidades reguladoras**, muy importantes en el ámbito Big Data en los aspectos de privacidad y legalidad relativa al uso de los datos.
- **Entidades de estandarización** nacionales e internacionales, de jure y de facto, cuyo objetivo será impulsar nuevos conceptos, sistemas y soluciones para adoptar globalmente estándares internacionales.
- **Redes colaborativas/ecosistemas**, donde diferentes actores de la cadena de valor colaboran para enriquecer sus servicios de valor añadido a sus clientes basados en los datos.

Asimismo, las diferentes tipologías de organizaciones también juegan uno o varios roles dentro de la cadena de valor:

- *Grandes empresas*, tradicionalmente proveedores de tecnologías y servicios aunque también pueden ser usuarias.
- *PYME*, que proporcionan conocimiento específico en un determinado ámbito y soluciones innovadoras para conceptos específicos.
- *Universidades*, que investigan en nuevos algoritmos y tecnologías que pueden aplicarse en herramientas y plataformas.
- *Centros de investigación*, que investigan en nuevos algoritmos y metodologías y que definen nuevos casos de negocio.

## 2.2 Áreas científico-tecnológicas prioritarias

Adentrándonos en el plano científico-tecnológico, es interesante reflejar las **5 áreas técnicas prioritarias** identificadas por la BDVA [6], en base a consultas a expertos de diferentes sectores, donde se deben desarrollar proyectos de investigación y desarrollo en el entorno del Big Data: **Ingeniería y gestión de datos, Arquitecturas optimizadas para analítica de datos estáticos y dinámicos, Analítica avanzada con algoritmos predictivos y prescriptivos, Visualización y experiencia de usuario avanzada, y Mecanismos de privacidad y anonimización.**

### Ingeniería y gestión de datos

La recolección de datos está aumentando rápidamente, pero los métodos y herramientas para la gestión de los mismos no están evolucionando al mismo ritmo. La capacidad de definir claramente, interoperar, compartir abiertamente, acceder, transformar, enlazar, syndicar y gestionar los datos es un reto en la actualidad. Para ello es crucial disponer de **técnicas de metadatos y semánticas** que permitan estructurar los conjuntos de datos, anotarlos, documentar los procesos asociados y generar o syndicar información a los receptores.

En este contexto, se han identificado los siguientes retos de investigación:

*Tsunami de datos.* Debido al incremento del ritmo de generación de los datos previsto para los próximos años, entre otras cosas motivado por el despliegue masivo de los **sistemas ciberfísicos** (*CPS-Cyber Physical Systems*) y el **Internet de las Cosas**, se requerirán **nuevos medios de almacenamiento de datos** que incorporen técnicas novedosas de reducción así como **nuevas aplicaciones para su integración, búsqueda y análisis.**

- *Multilingüismo.* Aunque gran parte de los recursos están en inglés, los datos de cada país se recolectan en el idioma local. Se deben conseguir enfoques integrados de análisis de datos e inteligencia de negocio, para lo cual la barrera del lenguaje tiene que superarse.
- *Silos de datos.* En todos los sectores se encuentran conjuntos de datos aislados y fragmentados. Para aumentar el impacto de las aplicaciones Big Data, se debe avanzar en el desarrollo de **técnicas que permitan la integración y el acceso inteligente y sin fisuras a las diferentes fuentes de datos.** También se requiere una colaboración eficiente entre todos los *players* (por ejemplo, utilizando los mismos estándares o *frameworks*).
- *Interoperabilidad.* Los diferentes productores y consumidores de datos, incluso del mismo sector, difieren en cuanto a los medios de almacenamiento, comunicación y acceso a los datos, por lo que es urgente construir una capa de **interoperabilidad entre todos los sistemas**, aprovechando las técnicas de transformación e integración semántica.
- *Curación.* Se requieren **servicios y herramientas de limpieza, integración y curación de datos** que permitan a los usuarios extraer la parte valiosa de los mismos para posteriormente integrarlos y prepararlos para los procesos de análisis posteriores.
- *Digitalización de datos y anotación semántica.* Las áreas claves a desarrollar aquí son medios automáticos y colaborativos para la **descripción semántica del contenido de datos no estructurados** así como la **interoperabilidad.**
- *Repositorios de conocimiento.* Es necesario **almacenar y compartir anotaciones, transformaciones, ontologías y conocimiento** de tal forma que se puedan enlazar los datos de la industria con los de los ciudadanos sin requerir conocimiento técnico experto.

## Arquitecturas optimizadas para analítica de datos estáticos y dinámicos

En los próximos años, el Internet de las Cosas será uno de los impulsores más importantes del Big Data. Compañías como Ericsson, Alcatel y Nokia están jugando un papel fundamental en este campo y Europa está liderando la aplicación del IoT en sectores como automoción (BMW, Bosch), transporte (Alstom, CAF, Siemens), ciudades inteligentes, salud (Siemens, Philips) y el aeroespacial (Thales, Airbus, Rolls Royce).

El reto consiste en la **integración de tecnologías y protocolos IoT en plataformas e infraestructuras Big Data.** Una de las soluciones técnicas que han adoptado compañías como SAP y Software AG es el procesado en memoria principal, pero el verdadero potencial se tendrá cuando se consigan diseñar **soluciones de arquitectura genéricas y optimizadas** que permitan el uso de datos estáticos y dinámicos.

Los principales retos tecnológicos identificados en esta área son los siguientes:

- *Técnicas analíticas.* Para extraer conocimiento de los datos y desarrollar aplicaciones de soporte a la toma de decisiones es fundamental aplicar **técnicas analíticas complejas a escala y para datos dinámicos**. Por ejemplo, sistemas predictivos como los motores de recomendación deben ser capaces de proveer predicciones en tiempo real a la vez que enriquecen las bases de datos de histórico, haciendo uso de modelos estadísticos complejos y refinados.
- *Internet de las Cosas.* Se requieren **algoritmos para analítica de flujos de datos IoT**.
- *Rendimiento.* El **rendimiento de los algoritmos** debe ser escalado varios órdenes de magnitud a la vez que se debe velar por la **reducción del consumo energético**, haciendo importantes esfuerzos en la **integración del hardware y del software**.
- *Procesamiento de flujos de datos (analítica y procesado de flujos de datos en tiempo real).* Es crucial para un procesado de flujos de datos efectivo desarrollar técnicas, herramientas e infraestructuras en áreas como el **razonamiento inductivo (machine learning)**, el **razonamiento deductivo** (inferencia), la **computación de alto rendimiento** (optimización de centro de datos, distribución eficiente de recursos, provisión de calidad de servicio) y el **análisis estadístico**. También son necesarias **nuevas técnicas de paralelización** específicas para Big Data y **distribución automática de tareas sobre clústeres**.

## Analítica avanzada con algoritmos predictivos y prescriptivos

El **impacto más significativo del Big Data** se espera que venga de la evolución de la analítica, influenciando toda la cadena de valor y generando oportunidades de negocio así como beneficios sociales y para los ciudadanos. La analítica avanzada es un campo activo abierto donde Europa tiene ventajas competitivas, donde hay gran potencial para el desarrollo de negocio, y que proporciona un valor adicional en los sectores donde se aplica al **posibilitar procesos más eficientes y precisos**. Se considera uno de los principales impulsores del crecimiento económico europeo, por encima de la producción de gas y petróleo.

En este ámbito es necesario crear **modelos predictivos complejos y refinados sobre conjuntos de datos masivos y heterogéneos** (como series temporales o datos gráficos). Además, estos modelos deben ser **aplicados en tiempo real sobre grandes cantidades de datos en movimiento, tanto estructurados como no estructurados**, lo que requiere de escalabilidad y alto rendimiento.

La comprensión de los datos, sean números, texto o contenido multimedia, siempre ha sido uno de los retos de las TIC. En la era del Big Data este reto ha propiciado la necesidad de desarrollar nuevos métodos en las siguientes áreas tecnológicas:

- *Datos gráficos.* **Modelos predictivos, técnicas de aprendizaje avanzado y técnicas de minería de datos en gráficos extremadamente grandes.**
- *Contexto.* En el ámbito del Big Data, la contextualización mejora la calidad de los procesos de minería o de los clasificadores. De hecho, en los últimos años

se ha demostrado que el **modelado de contexto** es crucial para conseguir una gestión optimizada de recursos, sistemas y servicios en cualquier dominio de aplicación.

- *Validación de contenidos.* **Modelos de veracidad** (fiabilidad de la fuente versus credibilidad de la información) para validar el contenido y explotar las recomendaciones de contenido de usuarios anónimos.
- *Frameworks.* **Nuevos marcos de trabajo y APIs abiertas para la distribución de tareas de procesamiento de datos** entre los recursos del data center, que tengan en cuenta la calidad y simplificando el trabajo de los desarrolladores y expertos.
- *Procesamiento.* **Mejora de la escalabilidad y velocidad de procesamiento de los algoritmos** mencionados para alcanzar una optimización computacional.
- *Analítica de negocio.* Los puntos anteriores permiten analítica o inteligencia de negocio que favorece la **toma de decisiones precisas y en el momento justo**. Es necesario simplificar y automatizar este proceso, especialmente en las PYMES.

## Visualización y experiencia de usuario avanzada

La visualización de los datos es vital si se quiere un consumo efectivo de Big Data. Los informes que se generan de la analítica de datos pueden ser documentos complejos. Para que los usuarios interpreten rápida y correctamente toda esta cantidad de documentos, se deben diseñar cuidadosamente la visualización digital y la presentación. Las **interfaces de información deben "humanizarse"**, es decir, responder a las necesidades y capacidad humanas.

Las herramientas de visualización actuales deben mejorarse para afrontar los retos que suponen el Volumen y la Variedad del Big Data. **Las técnicas de visualización avanzada deben considerar esta cantidad y heterogeneidad de datos disponibles** (gráficos, geoespacial, sensores, móviles, etc.).

En cuanto al acceso a los datos, el modelo tradicional de "bajo petición del usuario" en el ámbito del Big Data ya no es el más eficiente, por contra, la información debe extraerse proactivamente a través de **técnicas de descubrimiento de datos anticipándose a las necesidades de información del usuario**.

Las principales innovaciones tecnológicas necesarias en este ámbito según la BDVA son:

- *Centrado en el usuario final.* Visualización y analítica adaptada a las necesidades de los usuarios, incluyendo la posibilidad de feedback.
- *Abstracción.* Para manejar los volúmenes extremadamente grandes de datos, la interacción debe focalizarse sobre datos agregados a diferentes niveles de abstracción, más que sobre objetos individuales.
- *Clústeres.* Clusterización dinámica de información basada en semejanzas o en relatividad respecto a un problema, más que respecto a categorías individuales.

- *Geoespacial*. Nueva visualización de datos con geolocalizaciones, distancias y correlaciones espacio-tiempo.
- *Datos interrelacionados*. Las interfaces visuales deben tener en cuenta relaciones espaciales y semánticas de los datos, como posiciones, distancias, correlaciones espacio/tiempo, gráficos y redes de datos.
- *Análisis cualitativo*. Los enfoques clásicos cuantitativos deben ser superados con análisis cualitativo a un nivel semántico alto.
- *Tiempo*. Se deben tener en cuenta las especificidades del tiempo, cuya estructura semántica inherente y su sistema jerárquico de granularidad debe ser considerado.
- *Plug and Play*.

## Mecanismos de privacidad y anonimización

La anonimización y privacidad de los datos es una de las mayores preocupaciones en el ámbito del Big Data y de la analítica de datos, e implica a todos los *players* de la cadena de valor. De hecho, la **privacidad y seguridad de los datos** es con frecuencia un obstáculo fundamental para que los dueños de los datos sean partícipes de los entornos de innovación en Big Data. A su vez también los ciudadanos están exigiendo cada vez más esta garantía de privacidad.

En este ámbito existen muchas áreas en investigación actualmente como son la privacidad diferencial, la extracción de información privada y la encriptación homomórfica, que, aunque parecen ser prometedoras, no están disponibles comercialmente a gran escala.

Un aspecto importante es la provisión de **herramientas que evalúen el riesgo asociado al procesado combinado de múltiples conjuntos de datos**, en lo que a privacidad se refiere. Con la integración de múltiples fuentes de datos, las oportunidades de romper el proceso de anonimización aumentan considerablemente. **Asegurar la irreversibilidad de la anonimización** es uno de los aspectos clave en Big Data.

En este contexto, algunas de las innovaciones tecnológicas necesarias son:

- Protección de la infraestructura cloud, de las aplicaciones de analítica y de los datos ante pérdidas y amenazas.
- Métodos para el borrado seguro de datos y para la minimización de datos.
- Algoritmos robustos de anonimización.
- Herramientas de evaluación de riesgos que permitan valorar la reversibilidad de los mecanismos de anonimización.
- Algoritmos de minería de datos que preserven la privacidad.
- Mecanismos para la publicación y computación de datos que preserven la privacidad.

- Mecanismos de encubrimiento de patrones.
- Mecanismos de minería multipartidaria segura sobre conjuntos de datos distribuidos.

## 2.3 Ecosistema de tecnologías existentes

En la actualidad, Big Data es la respuesta a una gran variedad de retos tecnológicos que no tienen una única respuesta. A pesar de que es un término popularmente asociado a la tecnología Hadoop<sup>6</sup>, cuando hablamos de Big Data debemos pensar en un **ecosistema de tecnologías** que son capaces de solucionar problemas desde diferentes tipos de enfoques, y que en ocasiones deben hacerlo mediante la suma de varias de ellas.

Por tanto, la arquitectura de nuestra solución Big Data pasa por confeccionar un mapa de tecnologías que, mediante la orquestación, serán capaces de resolver uno o varios tipos de problemas.

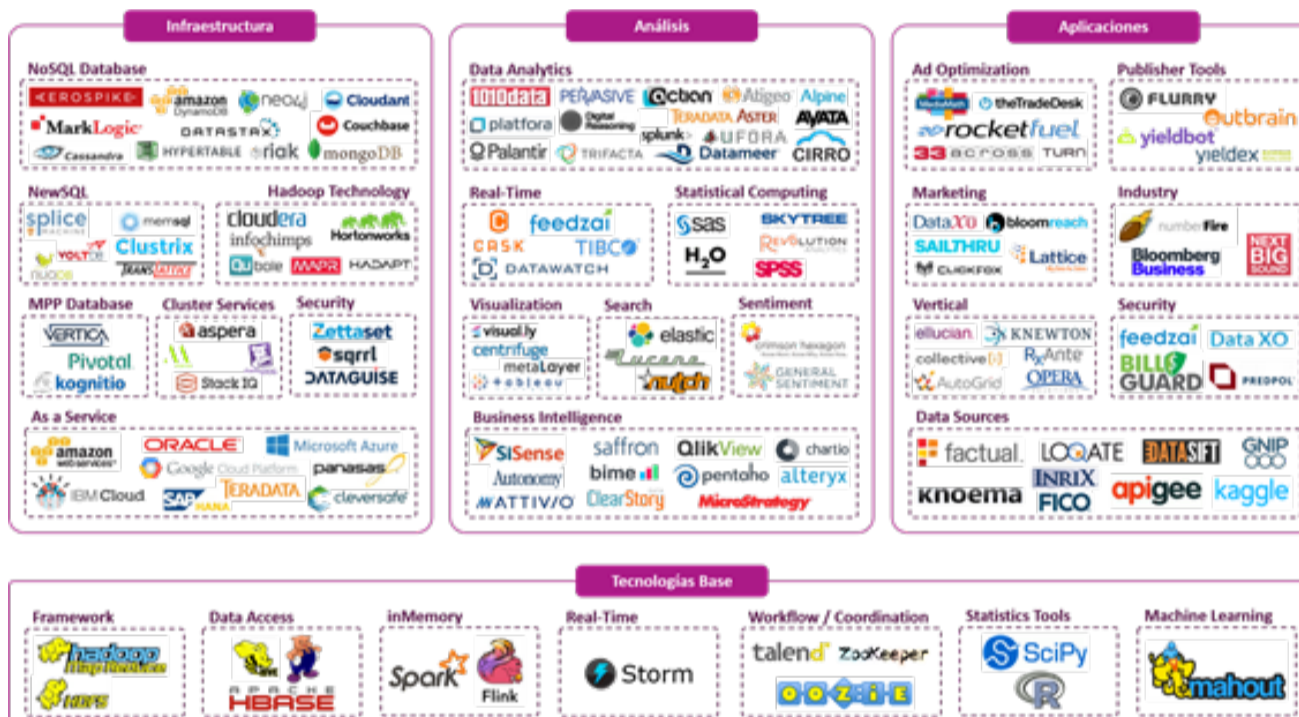


Figura 10. Tecnologías del Ecosistema de Big Data (Fuente: Elaboración Propia)

Las tecnologías Big Data deben permitir la construcción de un entorno que incluya el almacenamiento de datos, su procesamiento, el análisis, la visualización y los procesos de negocio. En este sentido, las tecnologías pueden agruparse en cuatro grandes categorías: *tecnología base, infraestructura, análisis y aplicaciones*.

### 2.3.1 Tecnología base

La mayoría de las soluciones y tecnologías Big Data que se desarrollan en la actualidad, parten desde dificultades similares a la hora de tratar grandes cantidades de datos, resolviendo retos relacionados con la captura, almacenamiento, análisis, visualización y apoyo a la toma de decisiones. Las **tecnologías base** Big Data se centran, principalmente, en proporcionar paliativos a estas dificultades de forma genérica, por lo que han servido de apoyo a la construcción de nuevas soluciones especializadas o más eficientes. Estas tecnologías base se apoyan en tres pilares:

6 Apache Hadoop es un framework que soporta el procesamiento de grandes volúmenes de datos a través de clústers, aplicando la filosofía MapReduce y el sistema de ficheros distribuidos escalable HDFS.

- Sistemas de archivos distribuidos: su objetivo principal es ofrecer alto rendimiento, escalabilidad y tolerancia a fallos para trabajar con infinidad de ficheros de manera simultánea. Principalmente, basado en el sistema **HDFS** (Hadoop Distributed File System) y su arquitectura Maestro-Esclavo.
- Bases de datos escalables: su objetivo es almacenar y procesar grandes volúmenes de datos con fiabilidad y bajos tiempos de respuesta. Como veremos en la categoría Infraestructura, se tratan de bases de datos **NoSQL** basadas en clave-valor o en *column-family*.
- Software de tratamiento masivo: su objetivo es conseguir repartir las necesidades computacionales para ejecutar un programa (realizar un cálculo, etc.) entre diversos nodos o servidores. Basado en el modelo **MapReduce** que representa una evolución del algoritmo de Divide y Vencerás (DYV) de computación distribuida.

El resultado ha sido la adopción mundial de **Hadoop** como framework base de la mayoría de soluciones actuales y que realiza funciones de middleware para la orquestación de ellas. Actualmente, se trata de un proyecto *open source* liderado por Apache que tiene multitud de variantes. Estas serían las tecnologías base en las que se apoya el ecosistema Big Data:

- **Framework:** Hadoop MapReduce, HDFS
- **Data Access:** Apache HBase, Apache Pig, Apache Hive, Cassandra, MongoDB
- **InMemory:** Apache Spark, Apache Flink
- **Workflow / Coordination:** Zookeeper, Talend, Oozie
- **Real-Time:** Storm
- **Statistical Tools:** R, SciPy
- **Machine Learning:** Apache Mahout

### 2.3.2 Infraestructura

Las tecnologías de infraestructura son los elementos más importantes del ecosistema Big Data, siendo las encargadas de procesar, almacenar y, en ocasiones, analizar o mover grandes volúmenes de datos. Durante décadas, las empresas han utilizado bases de datos relacionales para procesar información estructurada, pero éstas no han sido capaces de ofrecer un buen rendimiento para manejar grandes cantidades de datos complejos. El auge de los datos no estructurados ha supuesto ir más allá de meras filas y tablas, impulsando tecnologías de infraestructura capaces de manejar grandes Volúmenes de una gran Variedad de datos a una Velocidad muy competente, ejecutando aplicaciones en paralelo de sistemas con miles de nodos y de terabytes de datos.

Las principales características que definen este tipo de tecnologías de infraestructura son:

- No usan un modelo relacional, sino el denominado **NoSQL** para el acceso eficiente a datos no estructurados. Los mecanismos principales de almacenamiento que implementan son:

- **Almacén en Clave-Valor:** es la más simple de todas. Se crean pares clave-valor por cada entrada en la BBDD. La clave es única y el valor es en realidad un blob que puede incluir cualquier tipo de información. De hecho, la BBDD no tiene idea de lo que se almacena dentro del campo "valor". Eso es responsabilidad de la aplicación que explota los datos. Este mecanismo tan simple es el que permite que sean altamente escalables y que ofrezcan un rendimiento muy alto. Ejemplos de este tipo de soluciones son Riak, CouchBase, o DynamoDB de Amazon.
- **Almacén en Documento:** en este tipo de BBDD se almacenan documentos de tipo XML, JSON, BSON... Dentro del XML o el JSON, se especifican muchas características que describen el documento y permiten examinarlo y consultarlo. En realidad, sería un esquema clave-valor en el que el valor es un documento. MongoDB es, quizás, la más representativa.
- **Almacén en Columna:** es el sistema más parecido a las bases de datos relacionales. La información se almacena en filas cuyas columnas pueden ser diferentes entre ellas, permitiendo así tener información no estructurada. Además, es posible tener también columnas compuestas de columnas. Cassandra o HBase son alguno de los ejemplos más representativos de este tipo de solución.
- **Almacén en Grafos:** representan la información como nodos de un grafo y sus relaciones son utilizadas para recorrer la base de datos. Su utilidad se centra en representar las relaciones en el ámbito social, geográfico, software de recomendación y controles de acceso. Neo4j es la más utilizada.
- Están especialmente diseñadas para ejecutarse de forma distribuida o en **modo clúster**, permitiendo dividir y paralelizar cálculos complejos entre un número indefinido de nodos. Estas tecnologías se concentran principalmente en:
  - **NewSQL:** agrupa a las nuevas bases de datos relacionales y a los servicios asociados que están emergiendo para llevar los beneficios de un modelo relacional hacia una arquitectura distribuida y, en muchos casos, multi-tenant. Se trata de mejorar el rendimiento de las BBDD relacionales sin necesidad de escalar horizontalmente y ocultando y simplificando, dentro de lo posible, las tareas necesarias de partitioning, sharding, caching, o clustering. En realidad, se puede decir que el término que viene a agrupar soluciones que los proveedores de sistemas de BBDD relacionales ya venían desarrollando, pero para las que no había aún una categoría en la que incluirse.
  - **Massively Parallel Processing (MPP) Databases:** bases de datos SQL de trabajo mediante la segmentación de datos a través de múltiples nodos y su procesamiento en paralelo. Mientras Hadoop se suele ejecutar en racimos más baratos de los servidores de las materias primas, la mayoría de las bases de datos del MPP se ejecutan en un costoso hardware especializado.

Debemos tener en cuenta que las tecnologías Big Data especializadas en infraestructura han evolucionado para ofrecer servicios de valor añadido en otros aspectos como el análisis, por lo que resulta complejo ubicarlas en alguna subcategoría en concreto:

- **NoSQL Database:** Aerospike, Cassandra, CouchBase, Amazon Dynamo, IBM Cloudant, Marklogic, MongoDB, Riak, Neo4j, Datastax, Hypertable.
- **NewSQL Database:** Clustrix, MemSQL, VoltDB, NuoDB, SpliceMachine, Translattice.
- **MPP Database:** Vertica, Kognitio, Pivotal
- **Hadoop Technologies:** Cloudera, Hortonworks, MapR, Pivotal, Hadapt, Infochimps, QuBole
- **Cluster Services:** Datadog, StackIQ, Aspera, Mortar
- **As a Service:** Amazon EC2, Microsoft Azure, IBM Cloud, Oracle Cloud, Google Cloud Platform, Panasas, Cleversafe, Teradata, SAP Hana
- **Security:** Zettaset, Sqrrl, Dataguise

### 2.3.3 Análisis

*Big Data Analytics* hace referencia al proceso de recoger, organizar y analizar grandes volúmenes de datos para obtener información de valor. Aunque algunas de las tecnologías vistas en *Infraestructura* incorporan elementos para el análisis de datos, podemos encontrar otras que están específicamente diseñadas para este proceso. Por esta razón, las tecnologías identificadas en esta categoría participan en alguna de las siguientes subcategorías:

- **Analytics:** permiten integrar y analizar datos para descubrir nuevos conocimientos o realizar predicciones. Para ello, se implementan diversas técnicas de análisis de datos que incluyen la asociación de datos, la minería, la agrupación y el análisis de texto, análisis de sentimientos, tiempo real, búsqueda, localización, etc.
- **Visualization:** diseñadas específicamente para la representación visual y multidimensional que permita obtener información de valor a partir de relaciones complejas. Para ello, podemos encontrar dashboards, infografías, gráficas, mapas, diagramas, etc.
- **Business Intelligence (BI):** utilizadas para integrar y analizar datos específicamente para implementar servicios tales como los informes de inteligencia de negocio, cuadros de mando y gráficas para el apoyo en los procesos de toma de decisiones.

Las tecnologías Big Data que están especializadas en estas y otras técnicas relacionadas con el análisis, son las siguientes:

- **Data Analytics:** Palantir, Platfora, Datameer, Pervasive, Digital Reasoning, 1010Data, Alpine, Atigeo, Ayata, Cirro, Ufora, Trifacta, Terradata Aster, Actian, Splunk
- **Data Visualization:** Centrifuge, Metalayer, Visual.ly, Tableau
- **Statistical Computing:** Sas, Revolution Analytics, Skytree, SPSS, H2O

- **Sentiment Analysis:** Crimson Hexagon, General Sentiment
- **Search:** Elastic, Apache Lucene, Apache Nutch
- **Real-time:** DataWatch, Feedzai, Cask, Tibco
- **Business Intelligence:** Alteryx, Clear Story, Qlikview, Sisense, Saffron, Autonomy, Pentaho, Chartio, Microstrategy, Bime, Attivio

### 2.3.4 Aplicaciones

Esta categoría hace referencia a todas las tecnologías generadas por empresas del Big Data o aquellas que giran en torno al mercado de la toma de datos para su análisis posterior. El resultado son soluciones diseñadas para usuarios finales que buscan la optimización en algún campo de aplicación concreto, incluyendo:

- **Specific Applications:** soluciones sectoriales diseñadas específicamente para resolver una problemática concreta. Como por ejemplo, en el ámbito de la Salud, el apoyo a la elaboración pronósticos médicos a través del contraste de millones de imágenes 3D, o en el ámbito de la Energía, reducir el consumo de energía mediante el análisis y la predicción del gasto de los sistemas.
- **Data Sources:** soluciones diseñadas para recoger, catalogar y servir grandes volúmenes de datos generados por multitud de fuentes como sensores industriales y uso común, redes sociales, base de datos públicas, etc.

El conjunto de soluciones y tecnologías en el dominio de Big Data con un enfoque al usuario final, están subcategorizadas de la siguiente forma:

- **Ad Optimization:** RocketFuel, Turn, The Trade Desk, 33Across, MediaMath
- **Publisher Tools:** Outhbrain, Yieldex, Yieldbot, Flurry
- **Marketing:** Lattice, ClickFox, BloomReach, Sailthru, DataXU
- **Industry:** The Next Big Sound, Bloomberg Business, , numberFire,
- **Vertical:** Collectivei, AutoGrid, Knewton, Ellucian, RxAnte, Opera Solutions
- **Security:** BillGuard, DataXO, PredPol, Feedzai
- **Data Sources:** Apigee, DataSift, Fico, Inrix, Konema, Loqate, Factual, Gnip, Kaggle

## 3

## AGENTES E INICIATIVAS RELEVANTES

Esta sección recoge una primera aproximación de **análisis externo** de Big Data con el objetivo de reflejar los principales actores y sus proyectos en dicho ámbito.

### 3.1 Proveedores

Este apartado se centra en ofrecer una visión general sobre las empresas que están liderando el mercado del Big Data e intentar trazar un perfil de sus principales actores.

En cuanto al mercado mundial del Big Data medido como el **retorno económico** que obtienen los **proveedores de hardware, software y servicios TIC** en dicho ámbito, según IDC<sup>5</sup> alcanzará **en 2018 los 41.520 millones de dólares**, siendo la **tasa de crecimiento anual en el periodo 2014-2018 del 26,24%**.

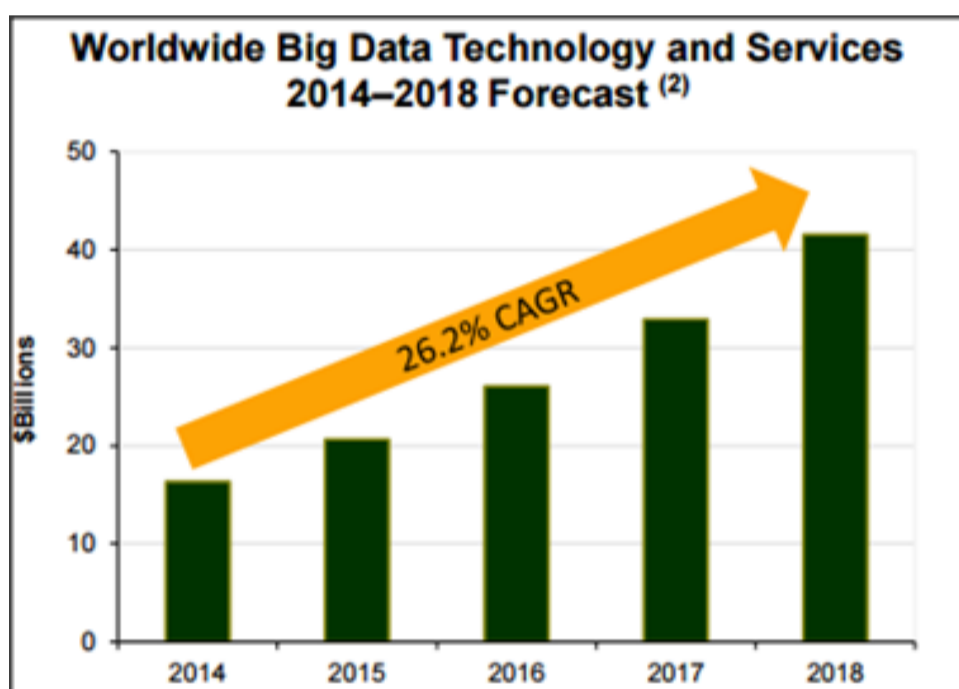


Figura 11. Worldwide Big Data Technology and Services 2014-2018 Forecast (Fuente: IDC)

Por su parte, la consultora tecnológica Wikibon<sup>6</sup> señala que el mercado de Big Data alcanzó en 2013 un volumen total de 18.600 millones de dólares, lo que supuso una tasa de crecimiento de un 58% respecto del año anterior. En dicho análisis se estima que **en 2017 el volumen de negocio podría superar los 50.000 millones de dólares**, marcando una tendencia claramente alcista de crecimiento anual a dos dígitos, fruto de la madurez y la mayor adopción de las tecnologías Big Data por parte de las empresas.

5 Worldwide Big Data Technology and Services 2014-2018 Forecast <http://www.idc.com/getdoc.jsp?containerId=250458>

6 Big Data Vendor Revenue and Market Forecast 2013-2017 [http://wikibon.org/wiki/v/Big\\_Data\\_Vendor\\_Revenue\\_and\\_Market\\_Forecast\\_2013-2017](http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017)

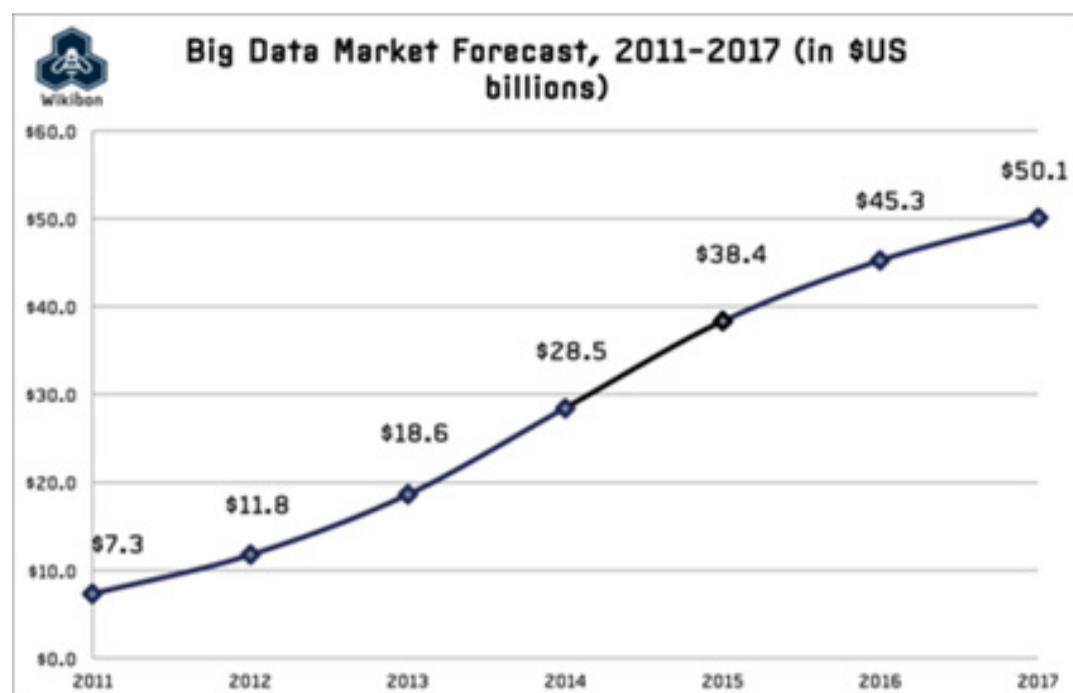


Figura 12. Big Data Vendor Revenue and Market Forecast 2013-2017 (Fuente: Wikibon 2014)

Desglosado por tipos, durante el ejercicio de 2013 los ingresos por servicios de Big Data supusieron el 40% del total del mercado, seguido por la venta de hardware con un 38% y la de software con un 22%. Este reparto en el negocio se explica por la circunstancia de que la mayoría de soluciones de software de este mercado y los modelos de negocio asociados de sus proveedores están relacionados con el código abierto. Además, existe la necesidad de ofrecer servicios profesionales que ayuden a las empresas a identificar casos de uso de Big Data, así como las arquitecturas e infraestructuras necesarias para poner en marcha proyectos de esta tipología.

La pregunta que cabe hacerse dadas estas cifras es: ¿qué **empresas o proveedores están liderando el mercado del Big Data**? Según los datos calculados por Wikibon, el reparto de los ingresos por Big Data entre empresas (proveedores) durante el 2013 queda de la siguiente manera:

Ingresos del mercado Big Data en 2014 a nivel mundial por proveedor (en millones de \$ americanos)							
Ranking	Proveedor	Ingresos por Big Data	Ingresos totales	% Ingresos Big Data respecto del total	% Ingresos Big Data Hardware	% Ingresos Big Data Software	% Ingresos servicios Big Data
1	IBM	\$1,368	\$99,751	1%	31%	27%	42%
2	HP	\$869	\$114,100	1%	42%	14%	44%
3	Dell	\$652	\$54,550	1%	85%	0%	15%
4	SAP	\$545	\$22,900	2%	0%	76%	24%
5	Teradata	\$518	\$2,665	19%	36%	30%	34%
6	Oracle	\$491	\$37,552	1%	28%	37%	36%
7	SAS Institute	\$480	\$3,020	16%	0%	68%	32%
8	Palantir	\$418	\$418	100%	0%	50%	50%
9	Accenture	\$415	\$30,606	1%	0%	0%	100%
10	PWC	\$312	\$32,580	1%	0%	0%	100%
11	Deloitte	\$305	\$33,050	1%	0%	0%	100%
12	Pivotal	\$300	\$300	100%	15%	50%	35%
13	Cisco Systems	\$295	\$50,200	1%	72%	12%	16%
14	Splunk	\$283	\$283	100%	0%	71%	29%
15	Microsoft	\$280	\$83,200	0%	0%	63%	37%
16	Amazon	\$275	\$70,000	1%	0%	0%	100%
17	Hitachi	\$260	\$89,999	1%	0%	0%	100%
18	CSC	\$188	\$14,200	1%	0%	0%	100%

19	CenturyLink	\$175	\$13,757	1%	0%	0%	100%
20	Google	\$175	\$59,767	1%	0%	0%	100%
21	Fusion-io	\$173	\$401	43%	90%	0%	10%
22	NetApp	\$167	\$6,450	3%	73%	0%	27%
23	Intel	\$165	\$52,708	1%	66%	21%	13%
24	EMC	\$165	\$23,222	1%	74%	0%	26%
25	Mu Sigma	\$160	\$160	100%	0%	0%	100%
26	TCS	\$157	\$11,570	1%	0%	0%	100%
27	Microstrategy	\$144	\$576	25%	0%	68%	32%
28	Actian	\$138	\$138	100%	0%	73%	27%
29	Booz Allen Hamilton	\$125	\$5,850	2%	0%	0%	100%
30	Opera Solutions	\$124	\$124	100%	0%	0%	100%
31	Red Hat	\$109	\$1,437	8%	0%	78%	22%
32	Capgemini	\$104	\$13,639	1%	0%	0%	100%
33	Informatica	\$98	\$948	10%	0%	82%	18%
34	MarkLogic	\$96	\$96	100%	0%	79%	21%
35	General Electric	\$80	\$146,000	1%	0%	75%	25%
36	VMware	\$80	\$5,207	1%	0%	79%	21%
37	Syncsort	\$75	\$75	100%	0%	97%	3%
38	Cloudera	\$73	\$73	100%	0%	53%	47%
39	SGI	\$65	\$667	10%	85%	0%	15%
40	MongoDB	\$62	\$62	100%	0%	56%	44%
41	Hortonworks	\$55	\$55	100%	0%	73%	27%
42	DDN	\$54	\$315	17%	81%	0%	19%
43	Guavus	\$54	\$54	100%	0%	61%	39%
44	Alteryx	\$48	\$48	100%	0%	90%	10%
45	1010data	\$45	\$45	100%	0%	0%	100%
46	Rackspace	\$42	\$1,520	3%	0%	0%	100%
47	TIBCO	\$36	\$1,069	3%	0%	64%	36%
48	MapR	\$35	\$35	100%	0%	77%	23%
49	Tableau Software	\$33	\$206	16%	0%	76%	24%
50	Qlik	\$30	\$467	6%	0%	73%	27%
51	Attivio	\$29	\$29	100%	0%	62%	38%
52	Juniper	\$28	\$4,669	1%	82%	0%	18%
53	DataStax	\$26	\$26	100%	0%	85%	15%
54	GoodData	\$26	\$78	33%	0%	0%	100%
55	Attunity	\$23	\$30	77%	0%	74%	26%
56	Fractal Analytics	\$19	\$27	70%	0%	0%	100%
57	Pentaho	\$18	\$38	45%	0%	76%	24%
58	Datameer	\$17	\$17	100%	0%	82%	18%
59	Couchbase	\$17	\$17	100%	0%	71%	29%
60	Basho	\$16	\$16	100%	0%	76%	24%
61	Kognitio	\$15	\$15	100%	0%	47%	53%
62	Sumo Logic	\$14	\$14	100%	0%	0%	100%
63	Jaspersoft	\$14	\$34	41%	0%	64%	36%
64	SiSense	\$14	\$14	100%	0%	79%	21%
65	Talend	\$14	\$57	25%	71%	0%	29%
66	Actuate	\$13	\$140	9%	0%	69%	31%
67	Revolution Analytics	\$12	\$12	100%	0%	67%	33%
68	Aerospike	\$12	\$12	100%	0%	92%	8%
69	Neo Technologies	\$12	\$12	100%	0%	67%	33%
70	Digital Reasoning	\$11	\$11	100%	0%	64%	36%
71	Tresata	\$10	\$10	100%	0%	90%	10%
72	Rainstor	\$10	\$10	100%	0%	70%	30%
73	Think Big Analytics	\$10	\$10	100%	0%	0%	100%
74	ODM	\$3,800	n/a	n/a	100%	0%	0%
75	Other	\$3,030	n/a	n/a	27%	20%	53%
	TOTAL	\$18,607	n/a	n/a	38%	22%	40%

Tabla 1. 2013 Worldwide Big Data Revenue by Vendor (\$US millions) (Fuente: Wikibon)

IBM es la líder destacada en la tabla, junto a HP, Dell y SAP. Entre las empresas líderes, ninguna de ellas está enfocada totalmente en este tipo de tecnologías, el porcentaje de ingresos que supone el Big Data en su negocio es totalmente residual (1-2%). Hay que bajar hasta la posición nº 8 (Palantir) para encontrar al primer proveedor Big Data "pure play", es decir, aquel que tiene en estas tecnologías su principal o única fuente de ingresos. Es interesante también resaltar el importante papel que están jugando consultoras internacionales como Deloitte, PWC o Accenture, situadas en las primeras posiciones de la tabla y teniendo como único tipo de ingreso en el sector del Big Data la venta de servicios derivados.

### 3.2 Plataformas tecnológicas y entidades asociadas

Las Plataformas Tecnológicas son redes o **estructuras colaborativas** formadas por agentes **privados y públicos**, que incluyen a todos los **actores relevantes en un campo tecnológico** particular (investigadores y científicos, empresas, autoridades públicas, etc.), pero que son **lideradas por la industria**, tanto en su creación como en la definición de sus modelos organizativos así como en la búsqueda de los instrumentos adecuados para su financiación.

La colaboración en plataformas tecnológicas supone muchas ventajas para las entidades participantes que redundan en una mejora de su competitividad. Las plataformas tecnológicas constituyen un foro de vigilancia tecnológica que proporciona **información de primera mano acerca del estado de la tecnología** en cada una de las áreas de actividad de un sector. Permiten el **traslado de los intereses** de los participantes de una manera eficaz **a las Administraciones Públicas**, facilitando la integración de sus **preferencias en programas de trabajo de I+D+i y en documentos de ámbito nacional y europeo**. También facilitan el **establecimiento de alianzas con otras organizaciones con iguales intereses**, fortaleciéndose las **oportunidades de negocio en cooperación**. Al mismo tiempo, representan un canal excelente para la **internacionalización** y, en definitiva, para **darse a conocer y dar visibilidad a la oferta tecnológica** de cada entidad al resto de actores relevantes en un ámbito tecnológico. Por último, permiten **reflejar el grado de interés de una región o país en un determinado ámbito**.

A continuación se destacan aquellas plataformas tecnológicas relacionadas con Big Data:

Actualmente, el principal foco de acción a nivel europeo sobre Big Data lo constituye la **Big Data Value Association (BDVA)** (<http://www.bdva.eu>), que nace de la **Plataforma Tecnológica Europea NESSI** (<http://www.nessi-europe.eu>) y que en octubre de 2014 firmó un compromiso con la Comisión Europea para la creación de una **Public Private Partnership (PPP) sobre Big Data**, en la que la Comisión va a poner más de 500 millones de euros entre 2016 y 2020 y donde las empresas se comprometen a realizar actividades por más de 1500 millones de euros adicionales.



*Figura 13. Lista de miembros fundadores de la Big Data Value Association*

A nivel Español, la **plataforma PLANETIC** (<http://planetec.es>), *mirror* de NESSI, ha puesto en marcha un grupo de trabajo inter-plataformas centrado en Big Data (**Iniciativa Big Data**<sup>7</sup>), cuyo objetivo es la colaboración entre distintas plataformas tecnológicas Españolas en las que el uso del Big Data puede suponer un impacto importante, como son:

- Plataforma de Medicamentos Innovadores (<http://www.medicamentos-innovadores.org>)
- Plataforma de Tecnología Sanitaria (<http://www.plataformatecnologiasanitaria.es>)
- Plataforma Food4Life (<http://www.foodforlife-spain.es>)
- Plataforma eNEM (<http://enem.imasdtic.es>)
- Plataforma es.INTERNET (<http://esinternet.imasdtic.es>)
- Plataforma Española del Agua (<http://www.plataformaagua.org>)
- Plataforma ManuKET (<http://www.manufacturing-ket.com/manu-ket/>)
- Plataforma Logistop (<http://logistop.org>)
- Plataforma de Tecnologías ambientales (<http://www.pt-planeta.es>)

En este marco, el pasado 16 de marzo PLANETIC organizó una jornada sobre Big Data en la sede del CDTI en Madrid, en el marco de su II Asamblea General, a la que asistieron 120 personas y contó con el apoyo del MINECO, CDTI, MINETUR y la PPP BigDataValue, contando con la presencia de varias de las plataformas tecnológicas implicadas en la iniciativa Big Data<sup>8</sup>.

Además del tratamiento de datos, el procesamiento de los mismos es un reto importante. Es por ello que la computación de altas prestaciones (HPC) cobra especial importancia en este entorno. La **plataforma ETP4HPC** (<http://www.etp4hpc.eu>) está trabajando activamente en colaboración con las plataformas de Big Data para generar sinergias y contribuir a la creación

7 Iniciativa Big Data <http://planetec.es/iniciativa-big-data-2015>; <http://planetec.es/sites/default/planetecfiles/content-files/page/06.%20Iniciativa%20Big%20Data%20-%20Clara%20Pezuela.pdf>

8 Jornada Big Data - II Asamblea General <http://planetec.es/jornadaBigData>

de las infraestructuras necesarias para soportar el manejo de grandes cantidades de datos.

Otras plataformas muy relacionadas con Big Data y que van a jugar un papel importante en alguno de los eslabones de la cadena de valor son:

### **Energía**

- EU PV TP ([www.eupvplatform.org](http://www.eupvplatform.org))
- TPWind ([www.windplatform.eu](http://www.windplatform.eu))
- SmartGrids ([www.smartgrids.eu](http://www.smartgrids.eu))

### **Transporte**

- ACARE ([www.acare4europe.com](http://www.acare4europe.com))
- ERRAC ([www.errac.org](http://www.errac.org))
- ERTRAC ([www.ertrac.org](http://www.ertrac.org))
- Logistics ([www.etp-logistics.eu](http://www.etp-logistics.eu))
- Waterbone ([www.waterborne-tp.org](http://www.waterborne-tp.org))

### **Producción y procesos**

- ECTP ([www.ectp.org](http://www.ectp.org))
- ESTEP ([cordis.europa.eu.int/estep/home\\_en.htm](http://cordis.europa.eu.int/estep/home_en.htm))
- EuMaT ([eumat.eu](http://eumat.eu)), FTC ([www.textileplatform.eu](http://www.textileplatform.eu))
- SusChem ([www.suschem.org](http://www.suschem.org))
- Nanomedicine ([www.etp-nanomedicine.eu](http://www.etp-nanomedicine.eu))
- ETP-SMR ([www.etpsmr.org](http://www.etpsmr.org))
- Manufuture ([www.manufuture.org](http://www.manufuture.org))
- Medioambiente
- WssTP ([wsstp.eu](http://wsstp.eu))

### **TIC**

- NEM ([www.neminitiative.org](http://www.neminitiative.org))
- NetWorld2020 ([www.networld2020.eu](http://www.networld2020.eu))

- ARTEMIS ([www.artemis-ju.eu](http://www.artemis-ju.eu))
- ENIAC ([www.eniac.eu](http://www.eniac.eu))
- EPoSS ([www.smart-systems-integration.org](http://www.smart-systems-integration.org))
- Photonics21 ([www.photonics21.org](http://www.photonics21.org))

Y por último, se han generado recientemente otras Agrupaciones Público Privadas (PPP) con enormes sinergias con las infraestructuras y el soporte al manejo de los datos:

- **FI-PPP** ([www.fi-ppp.eu](http://www.fi-ppp.eu)): PPP de Internet del Futuro, en la que se desarrollan infraestructuras que soportan el procesamiento de los datos.
- **5G-PPP** (<http://5g-ppp.eu/>): PPP de comunicaciones centrada en el desarrollo de infraestructuras de comunicaciones avanzadas.
- **FoF-PPP** ([http://ec.europa.eu/research/industrial\\_technologies/factories-of-the-future\\_en.html](http://ec.europa.eu/research/industrial_technologies/factories-of-the-future_en.html)): PPP centrada en Fabricación del Futuro, en la que el manejo de grandes cantidades de datos se configura como un pilar fundamental para el mantenimiento predictivo, la fabricación personalizada,...
- **HPC-PPP** (<http://www.etp4hpc.eu/hpc/what-is-hpc/>): Basada en el trabajo de la plataforma ETP4HPC.

Es interesante destacar también algunas iniciativas modélicas ya en marcha como son SDIL en Alemania y TeraLab en Francia.

**SDIL (Smart Data Innovation Lab)** (<http://www.sdil.de>) es una infraestructura de alto rendimiento destinada a potenciar la investigación en Big Data, a la vez que se reduce la brecha que hay entre la investigación y la aplicación de tecnologías de la información en la industria.

Esta plataforma está siendo operada por el Karlsruhe Institute of Technology (KIT). Además de KIT, entre sus fundadores se encuentran compañías como Bayer, Bosch, Microsoft Deutschland, SAP, Siemens y Software AG, así como el centro alemán de investigación en inteligencia artificial (DFKI), el Fraunhofer-Gesellschaft y Forschungszentrum Jülich. Cuenta así mismo con el apoyo de otras 20 compañías e instituciones como Infineon, Trumpf, Volkswagen, la asociación de industria TIC Bitkom y la sociedad de informáticos de Alemania (Gesellschaft für Informatik, GI).

Los primeros proyectos se focalizan en 4 áreas de investigación estratégicas: *industria 4.0, energía, ciudades inteligentes y medicina personalizada*.

SIDIL ofrece acceso a una gran variedad de tecnologías Big Data de última generación, tal y como se observa en la siguiente figura. En primera instancia ofrece software como SAP HANA, Software AG Terracotta e IBM Watson, y hardware como IBM Power e Intel architecture.

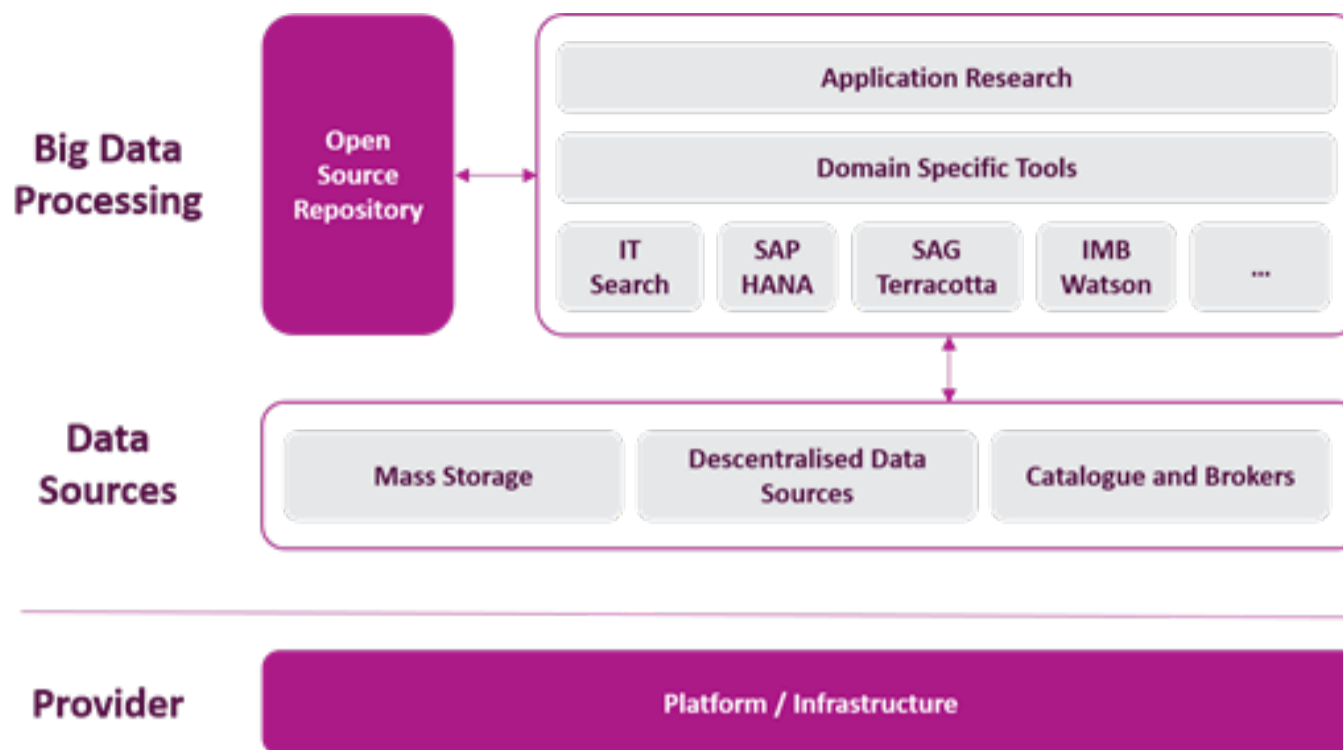


Figura 14. Smart Data Innovation Lab Structure

Los grupos de trabajo de cada área estratégica cuentan con representantes del lado de la ciencia y de la industria, y juntos deciden el tipo y la distribución de los recursos SDIL para los proyectos de investigación. En el marco de estos proyectos específicos, los científicos pueden acceder a datos de los procesos industriales que se almacenan de forma segura en la plataforma. El análisis de estos datos se hace de forma colaborativa con la empresa industrial, asegurando una rápida transferencia de conocimiento y tecnología. Además, los institutos de investigación implicados están desarrollando herramientas y métodos universales para el análisis de datos, que serán accesibles a los *partners* de SDIL.

**TeraLab** (<https://www.teralab-datascience.fr/fr/>) es una plataforma de servicios destinada a proyectos de investigación, innovación y enseñanza con relación a los datos masivos. Fue ganadora de la convocatoria de proyectos Big Data 2012 del Programa de Inversiones y Futuro (PIA), y lanzada oficialmente en febrero 2014.

Se trata de un entorno de investigación y experimentación para aplicaciones innovadoras o pilotos industriales, tanto para la comunidad de investigadores y docentes como para las empresas. Los promotores de proyectos disponen de esta forma de un entorno óptimo para consagrarse al tratamiento aplicativo de datos masivos, facilitar la producción y validar los resultados de investigación que pueden transferirse en innovaciones.

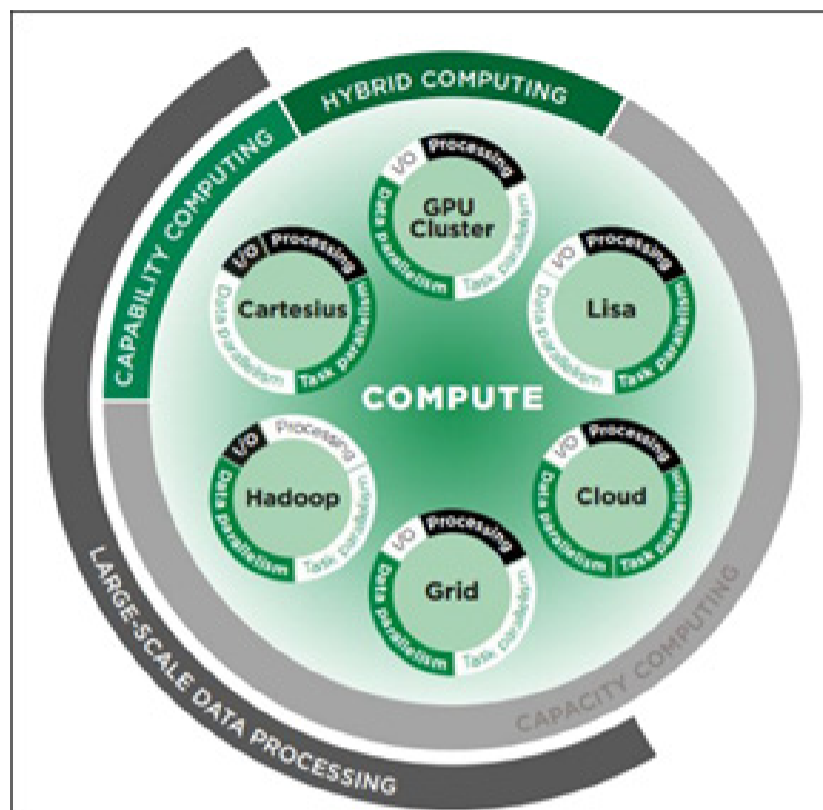
TeraLab está accesible para las PYMES, ETI y grandes grupos, al margen de cualquier explotación comercial, a través de proyectos colaborativos o bilaterales de tipo FUI, Eureka, Horizon 2020, EIT ICT Labs, ITEA, ANR, etc., para desarrollar innovaciones aplicativas y probar pilotos industriales.



*Figura 15. TeraLab Infrastructure*

La plataforma incluye medios materiales como una considerable capacidad de tratamiento con una memoria viva de varios terabytes, corpus de datos, aplicaciones y herramientas innovadoras. Está combinada con servicios que se basan en competencias especializadas específicas. La utilización se facilita en modo SaaS/PaaS a través de un catálogo de servicios y aplicaciones aportado por PYMES innovadoras y seleccionadas a través de un concurso.

En relación con las dos iniciativas anteriores y en el ámbito de las infraestructuras para potenciar la investigación de Big Data, destacar también a **SURFSARA** (<https://surfsara.nl/>), que desde hace años está ofreciendo servicios de infraestructura digital a la comunidad investigadora en Holanda, destacando sus capacidades de supercomputación (Cartesius) y los servicios de clúster computacional (Lisa), además de infraestructura Grid, infraestructura de Cloud Computing, un entorno de analítica Big Data (Hadoop), un servicio de almacenamiento de datos y un espacio de presentación y visualización avanzada.



*Figura 16. SURFSARA Infrastructure*

### 3.3 Proyectos I+D+I

Con el fin de conocer las iniciativas relevantes en materia de investigación, se han revisado los proyectos de Investigación y Desarrollo más recientes y destacados a nivel europeo en el ámbito de las TIC para Big Data.

Se ha elaborado una tabla con los datos básicos de cada proyecto más una relación de las entidades que forman el consorcio, acompañada de unas palabras clave que describen la investigación llevada a cabo. La fuente utilizada para extraer dicha información es CORDIS ([http://cordis.europa.eu/home\\_es.html](http://cordis.europa.eu/home_es.html)).

Esta información ha sido incluida con dos claros objetivos: por una parte, ofrecer una visión en conjunto de los actores europeos más activos en I+D para facilitar su análisis en una posible búsqueda de alianzas y, por otra, ofrecer de una manera sintética aquellos aspectos tecnológicamente más relevantes que permitan conocer las líneas de investigación llevadas a cabo actualmente de una manera rápida y directa.

En este último punto, destacar algunas de las materias TIC sobre las que se está investigando en este campo:

- Sistemas de análisis de información genómica
- Escalabilidad y eficiencia computacional de los sistemas Big Data
- Big Data para la industria y sistemas M2M
- Aspectos económicos, legales, éticos y políticos del Big Data

Algunos proyectos relevantes y recientes en el ámbito del Big Data son los siguientes:

- **BIG: Big Data Public Private Forum** (septiembre 2012 – octubre 2014, ICT-2011.4.4, <http://www.big-project.eu>). Acción de coordinación y soporte del VII PM cuyo objetivo es la definición e implementación de una estrategia clara en términos de investigación e innovación en Big Data a nivel Europeo.
- **BYTE: Big data roadmap and cross-disciplinary community for addressing societal Externalities** (marzo 2014 – febrero 2017, ICT-2013.4.2, <http://byte-project.eu/>). Acción de coordinación y soporte del VII PM cuyo objetivo es ayudar a la ciencia y a la industria europea en la captura de mayor mercado de Big Data en 2020. Para ello se llevarán a cabo una serie de casos de estudio en un conjunto de sectores industriales que permitirán aumentar el entendimiento de los aspectos económicos, legales, éticos y políticos del Big Data.
- **BIOBANK: Scalable, Secure Storage of Biobank Data** (diciembre 2012 – noviembre 2015, ICT-2011.4.4, <http://www.biobankcloud.com>). El proyecto tiene como objetivo el desarrollo de una plataforma en cloud computing (PaaS) para el almacenamiento, análisis e integración de información entre biobancos. La plataforma proveerá de herramientas de análisis y tecnologías Big Data que permitan analizar la enorme cantidad de datos derivados de la secuenciación genómica.
- **MR.SYMBIOMATH: High Performance, Cloud and Symbolic Computing in Big-Data problems applied to mathematical modeling of Comparative Genomics** (febrero 2013 – enero 2017, FP7-PEOPLE-2012-IAPP - Marie Curie Action, <http://www.mrsymbiomath.eu>). El objetivo del proyecto es trabajar de forma rápida y eficiente con el procesamiento de grandes conjuntos de datos informáticos (Big Data), procedentes de análisis y estudios genómicos, para su posterior aplicación en la detección precoz de alergias y reacciones a fármacos.
- **OPTIQUE - Scalable End-user Access to Big Data** (noviembre 2012 – octubre 2016, ICT-2011.4.4, <http://optique-project.eu>). El objetivo del proyecto es plantear un nuevo paradigma en cuanto al acceso y búsqueda de grandes cantidades de datos (Big Data) por parte de los usuarios finales para el apoyo efectivo de toma de decisiones críticas en las grandes empresas.
- **LeanBigData - Ultra-Scalable and Ultra-Efficient Integrated and Visual Big Data Analytics** (febrero 2014 – enero 2017, ICT-2013.4.2, <http://leanbigdata.eu>). El proyecto persigue la obtención de una plataforma big data ultra escalable y ultra eficiente capaz de aunar las capacidades de dos tipos de bases de datos —las operacionales y las analíticas— en una única, permitiendo un ahorro del 80% del coste en *business analytics*.
- **RETHINK big - Roadmap for European Technologies in Hardware and Networking for Big Data** (marzo 2014 – febrero 2016, ICT-2013.4.2, <http://www.rethinkbig-project.eu>). El objetivo del proyecto es aunar los intereses tanto de productores como de consumidores de macrodatos (*big data*), identificando las fortalezas de la industria europea con el fin de que adquiera en los próximos 10 años una posición de liderazgo en el ámbito del procesado y análisis.
- **FERARI - Flexible Event pRocessing for big dAtA aRchItectures** (febrero 2014 – enero 2017, ICT-2013.4.2, <http://www.ferari-project.eu>). El objetivo del proyecto es el desarrollo de tecnologías big data que permitan el procesamiento de los flujos de datos masivos en tiempo real generados por los sistemas M2M.

- **VELaSSCo - Visualization for Extremely Large-Scale Scientific Computing** (enero 2014 – diciembre 2016, ICT-2013.4.2, <http://www.velassco.eu>). El objetivo del proyecto es proporcionar herramientas de *Big Data* para la ingeniería y la comunidad científica que permitan manipular simulaciones con miles de millones de registros de una manera más fácil, detallada y rápida.
- **BigFoot - Big Data Analytics of Digital Footprints** (octubre 2012 – septiembre 2015, ICT-2011.1.2, <http://bigfootproject.eu>). El objetivo es diseñar, implementar y evaluar las maneras escalables para procesar e interactuar con aplicaciones de Big Data para la huella digital.

Acrónimo	Título. URL.	País (líder)		Duración	Palabras clave
BIG	Big Data Public Private Forum <a href="http://www.big-project.eu">http://www.big-project.eu</a>	España (ATOS SPAIN SA)	STI INTERNATIONAL CONSULTING UND RESEARCH GMBH UNIVERSITAET INNSBRUCK DEUTSCHES FORSCHUNGSZENTRUM FUER KUENSTLICHE INTELLIGENZ GMBH SIEMENS AKTIENGESELLSCHAFT INSTITUT FUR ANGEWANDTE INFORMATIK EV Open Knowledge Foundation Deutschland AGT GROUP (R&D) GMBH EXALEAD NATIONAL UNIVERSITY OF IRELAND, GALWAY The Press Association Ltd	From 2012-09-01 to 2014-10-31	Big Data Economy, Technology and strategy roadmap, Intelligent Information Management
BYTE	Big data roadmap and cross-disciplinary community for addressing societal Externalities <a href="http://byte-project.eu/">http://byte-project.eu/</a>	Reino Unido (TRILATERAL RESEARCH & CONSULTING LLP)	UNIVERSITAET INNSBRUCK VRIJE UNIVERSITEIT BRUSSEL SIEMENS AKTIENGESELLSCHAFT INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE NEMZETI INFORMACIOS INFRASTRUKTURA FEJLESZTESI IRODA NATIONAL UNIVERSITY OF IRELAND, GALWAY CONSIGLIO NAZIONALE DELLE RICERCHE TECHNISCHE UNIVERSITEIT DELFT DNV GLAS UNIVERSITETET I OSLO	From 2014-03-01 to 2017-02-28	Big data roadmap, big data community
BIOBANK	Scalable, Secure Storage of Biobank Data <a href="http://www.biobankcloud.com/">http://www.biobankcloud.com/</a>	Suecia (KUNGLIGA TEKNISKA HOEGSKOLEN)	CHARITE - UNIVERSITAETSMEDIZIN BERLIN HUMBOLDT-UNIVERSITAET ZU BERLIN FUNDACAO DA FACULDADE DE CIENCIAS DA UNIVERSIDADE DE LISBOA KAROLINSKA INSTITUTET FACULDADE DE CIENCIAS DA UNIVERSIDADE DE LISBOA	From 2012-12-01 to 2015-11-30	Genomics, DNA sequencing, biobank, data anonymization, PaaS, big data, cloud computing, health

MR.SYM-BIOMATH	High Performance, Cloud and Symbolic Computing in Big-Data problems applied to mathematical modeling of Comparative Genomics  <a href="http://www.mrsymbio-math.eu">http://www.mrsymbio-math.eu</a>	España (Universidad de Málaga)	RISC SOFTWARE GMBH UNIVERSITAT LINZ INTEGROMICS SL ITG Servicio Andaluz de Salud BAYERISCHE AKADEMIE DER WISSENSCHAFTEN	From 2013-02-01 to 2017-01-31	HPC, comparative genomics, bioinformatics, biomedical, health
OPTIQUE	Scalable End-user Access to Big Data  <a href="http://optique-project.eu">http://optique-project.eu</a>	Noruega (UNIVERSITETET I OSLO)	TECHNISCHE UNIVERSITAET HAMBURG-HARBURG FLUID OPERATIONS AG SIEMENS AKTIENGESELLSCHAFT NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS LIBERA UNIVERSITA DI BOLZANO UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA DNV GL AS STATOIL ASA THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD	From 2012-11-01 to 2016-10-31	Ontology, semantics, intuitive queries, mapping, data integration, RDSBMS
LEANBIG-DATA	Ultra-Scalable and Ultra-Efficient Integrated and Visual Big Data Analytics  <a href="http://leanbig-data.eu">http://leanbig-data.eu</a>	España (Universidad Politécnica de Madrid)	ATOS SPAIN SA CA TECHNOLOGIES DEVELOPMENT SPAIN SA FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS INTEL RESEARCH AND INNOVATION IRELAND LIMITED SYNC LAB SRL PT COMUNICACOES SA INESC PORTO - INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES DO PORTO	From 2014-02-01 to 2017-01-31	noSQL, CEP systems, scalable, efficient, analytics, workload
RETHINK big	Roadmap for European Technologies in Hardware and Networking for Big Data  <a href="http://www.rethinkbig-project.eu">http://www.rethinkbig-project.eu</a>	España (CENTRO NACIONAL DE SUPERCOMPUTACION)	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE TECHNISCHE UNIVERSITAET BERLIN PARSTREAM GmbH UNIVERSIDAD POLITECNICA DE MADRID ALCATEL - LUCENT BELL LABS FRANCE THALES SA NO RACK SAS INTERNET MEMORY RESEARCH SAS STICHTING CENTRUM VOOR WISKUNDE EN INFORMATICA ARM LIMITED THE UNIVERSITY OF MANCHESTER	From 2014-03-01 to 2016-02-29	Europe, hardware, networking, processing, algorithm, roadmap technology

FERARI	Flexible Event processing for big data architectures  <a href="http://www.ferari-project.eu">http://www.ferari-project.eu</a>	Alemania (FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V)	TECHNICAL UNIVERSITY OF CRETE POSLOVNA INTELIGENCIJA DOO ZA DIZAJN I IMPLEMENTACIJU INTELIGENTNIH INFORMACIJSKIH SUSTAVA HRVATSKI TELEKOM TECHNION ISRAEL INSTITUTE OF TECHNOLOGY IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD UNIVERSITY OF HAIFA	From 2014-02-01 to 2017-01-31	M2M, complex event processing, real time processing, telecommunications
VELaSSCo	Visualization for Extremely Large-Scale Scientific Computing  <a href="http://www.velassco.eu">http://www.velassco.eu</a>	España (CENTRE INTERNATIONAL DE METHODES NUMERICS EN INGENIERIA)	FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V ATOS SPAIN SA INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE STIFTELSEN SINTEF JOTNE EPM TECHNOLOGY AS THE UNIVERSITY OF EDINBURGH	From 2014-01-01 to 2016-12-31	Modeling applications, real time analysis, simulations, scalable algorithms, industry
BigFoot	Big Data Analytics of Digital Footprints  <a href="http://bigfoot-project.eu">http://bigfoot-project.eu</a>	Francia (EU-RECOM)	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE TECHNISCHE UNIVERSITAET BERLIN GRIDPOCKET SAS SYMANTEC LIMITED	From 2012-10-01 to 2015-09-30	digital footprint, parallel processing, private cloud, cross-layer optimization, open source

*Tabla 2. Proyectos de I+D+i internacionales relacionados con Big Data (Fuente: Elaboración propia)*

## DOMINIOS DE APLICACIÓN

### 4.1 Sectores principales e impactos esperados

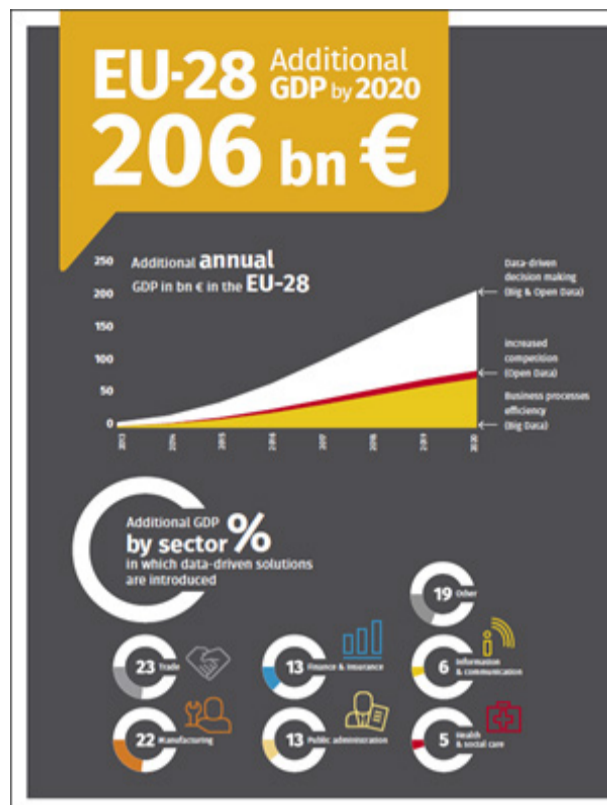
Son numerosos los sectores empresariales que ya se están beneficiando o podrían aprovecharse de las **ventajas competitivas** que aporta la extracción de conocimiento a partir de grandes cantidades de datos: **eficiencia, eficacia, productividad, calidad, valor añadido en los productos y servicios**, etc.

Por su parte, los **usuarios y ciudadanos también son beneficiarios directa e indirectamente** de la innovación que propicia el Big Data, por ejemplo, a través de sistemas y servicios de salud mejorados, de un mayor acercamiento y transparencia con las instituciones gubernamentales, de la personalización de los productos y servicios, etc.

A nivel cuantitativo, el análisis realizado por el Warsaw Institute for Economic Studies [3], tal y como se mostró en el apartado introductorio, indica que la transición a soluciones tecnológicas basadas en datos (*Big Data & Open Data*) supondrá una **aportación de 206.000 millones de euros para la economía de Europa en el año 2020**, repercutiendo en un **incremento medio del 1,9% del PIB de la Eurozona**, con especial impacto en el Norte de Europa.

Por sectores de aplicación, tal y como se muestra en la siguiente figura, dicho estudio indica un incremento del PIB de la Eurozona, debido a las soluciones basadas en datos, del:

- 23% para **Comercio**, lo que equivale aproximadamente a **47.000 millones de euros**.
- 22% para **Industria**, lo que equivale aproximadamente a **45.000 millones de euros**.
- 13% para **Administración Pública**, lo que equivale aproximadamente a **27.000 millones de euros**.
- 13% para **Financiero y aseguradoras**, lo que equivale aproximadamente a **27.000 millones de euros**.
- 6% para **TIC**, lo que equivale aproximadamente a **12 millones de euros**.
- 5% para **Socio-Sanitario**, lo que equivale aproximadamente a **10.000 millones de euros**.



*Figura 17. Incremento del PIB por sectores en Europa y en 2020 debido a soluciones basadas en datos (Fuente: WISE Institute, [3])*

Según el McKinsey Global Institute [7], la adopción de Big Data puede generar considerables retornos económicos en diferentes sectores tales como el sanitario, el sector público, la distribución, y la fabricación, entre otros. En la siguiente figura podemos ver las estimaciones realizadas por McKinsey en este sentido:

- Aportación de **300.000 millones de dólares al año** y crecimiento de la productividad del 0.7% anual, en el **sector socio-sanitario de Estados Unidos**.
- Aportación de **250.000 millones de euros al año** e incremento de la productividad del 0,5% anual, para la **administración pública europea**.
- Más de **100.000 millones de dólares de retorno para los proveedores de servicios** de datos y hasta **700.000 millones de dólares para los usuarios finales**, en el sector de la localización personal a nivel global.
- Más de un **60% de incremento en el margen neto** y un crecimiento de la productividad de entre el 0,5 y el 1% anual, en el **sector de la distribución de Estados Unidos**.
- **Una reducción de hasta el 50% en los costes de fabricación en el sector manufacturero**.



Figura 18. Potencial valor financiero generado por Big Data en diferentes sectores (Fuente: McKinsey [7])

Este estudio nos ilustra también cómo existen **determinados sectores mejor posicionados** para obtener retornos del Big Data, teniendo en cuenta el crecimiento de productividad que han tenido a lo largo de los últimos años en Estados Unidos:

- El **sector TIC** destaca como uno de los sectores que ha experimentado un fuerte crecimiento en productividad y que está posicionado para beneficiarse considerablemente del uso del Big Data.
- Los **sectores de servicios financieros y aseguradoras**, así como las **administraciones** están posicionados como grandes beneficiarios del Big Data una vez superen las barreras existentes, principalmente la resistencia a usar analíticas de Big Data.
- En un tercer nivel aparecen sectores que han experimentado crecimientos negativos como la construcción, los servicios educativos, las artes y entretenimiento, etc., donde aparecen fuertes barreras sistémicas.

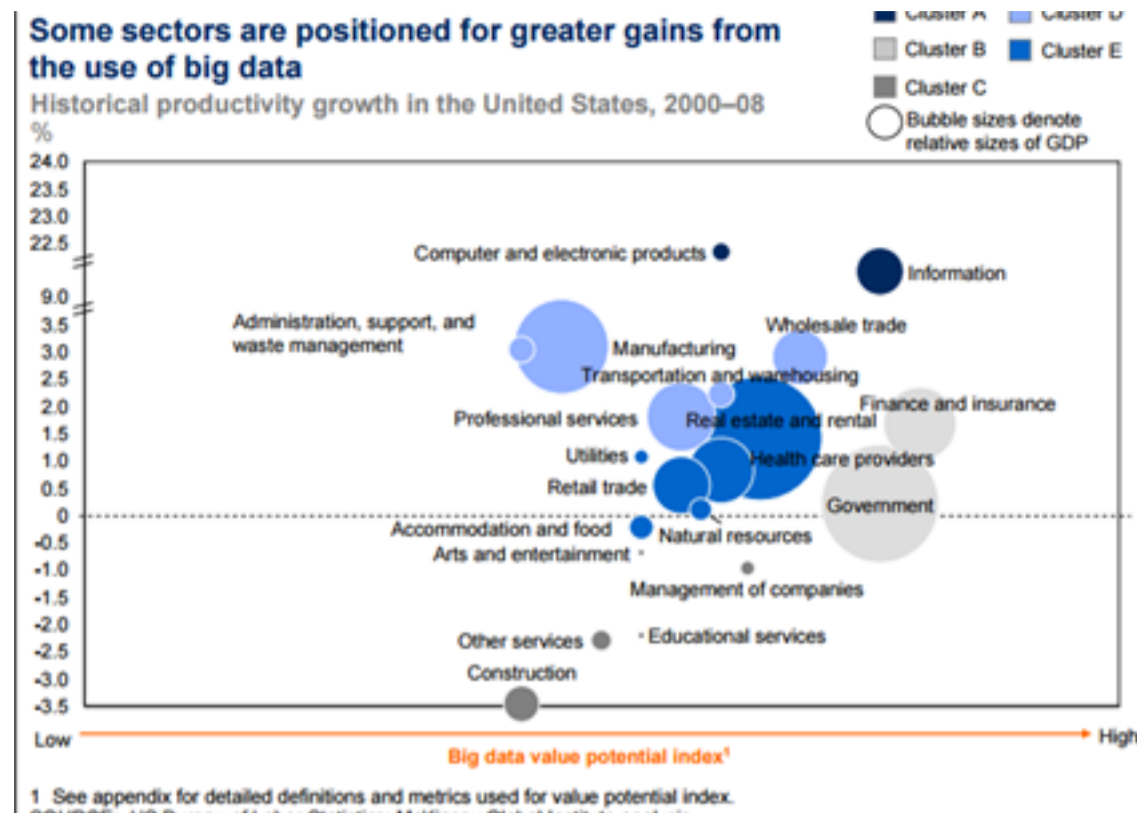


Figura 19. Posicionamiento de los sectores para obtener beneficio del Big Data (Fuente: Fuente: McKinsey [7])

Por su parte, el análisis realizado por la Big Data Value Association en su Agenda Estratégica de Investigación e Innovación [6], también muestra los impactos esperados del Big Data en diferentes sectores o dominios de aplicación:

- Hasta **300.000 millones de euros en la administración pública europea**, considerando los 23 estados más grandes en la Unión Europea.
- **90.000 millones de euros en el sector socio-sanitario europeo**, considerando únicamente la potencial reducción del gasto socio-sanitario en la Unión Europea.
- **79.000 millones de euros en el sector de las "utilities" global**, considerando una reducción de las emisiones de CO<sub>2</sub> de más de 2 giga toneladas.
- **500.000 millones de dólares en el sector de transporte y logística a nivel mundial**, considerando los ahorros de tiempo y gasolina, o la reducción de emisiones de CO<sub>2</sub> en 380 mega toneladas.
- **60% de incremento potencial en los márgenes de operación en el sector distribución y comercio.**
- **800.000 millones de dólares de retorno total en el sector de geolocalización global**, incluyendo los proveedores de servicio y los usuarios finales.
- **51.000 millones de dólares a nivel mundial directamente asociados al sector de servicios y aplicaciones Big Data.**

A continuación se analizan los principales sectores en los que el uso y aprovechamiento de la gestión y el análisis de datos masivos suponen ventajas competitivas nada despreciables:

## SALUD

Un campo importante de aplicación del Big Data es el de la salud y la atención sanitaria, posibilitando una **medicina "4P": personalizada, predictiva, preventiva y participativa**.

Tal y como se indica en el estudio del OBS, el uso de Big Data en este ámbito está siendo beneficioso en dos ejes principales. Por un lado, el **conocer mejor a los pacientes** (a través de sensores inteligentes, dispositivos móviles u otros mecanismos que registran su actividad) supone una mayor **eficiencia en el proceso de atención** y, por tanto, una **mayor calidad en el servicio** prestado y un **ahorro de costes**. Por el otro, los avances tecnológicos en cuanto a registro y procesamiento de grandes cantidades de datos han supuesto la **evolución de los procesos de análisis de información** sanitaria, dando lugar a los llamados **sistemas cognitivos**, pasando de conocer sólo lo que ha ocurrido (el pasado) a disponer de sistemas computacionales capaces de interpretar el significado y el contexto, **analizando y entendiendo enormes cantidades de datos en tiempo real**.

En este sentido, destacar que la mayoría de los datos en este sector son no estructurados, provienen de múltiples sistemas que normalmente no están debidamente conectados y no se tienen resultados consolidados. Además, la generación de información sanitaria está alcanzando ritmos muy fuertes.

Datos estructurados en el ámbito sanitario son los clásicos de los pacientes (nombre, edad, sexo...) y datos no estructurados son las recetas de papel, los registros médicos, las notas manuscritas de médicos y enfermeras, las grabaciones de voz, las radiografías, escáneres, resonancias magnéticas, TAC y otras imágenes médicas.

En cuanto al origen de los datos sanitarios, provienen de los más variados dispositivos, sensores, aparatos médicos diversos, datos hospitalarios...sumados a los datos procedentes de medios sociales (redes sociales, blogs, wikis...), de teléfonos inteligentes, etc.

Así, la incorporación de este tipo de sistemas cognitivos al campo de la medicina supone un **cambio radical en los procesos de diagnóstico o en la toma de decisiones, basado en evidencias**: *mayor eficacia en el diagnóstico, reducción al máximo de los errores a la hora de prescribir, entregar y pautar los medicamentos, y minimización de los errores médicos que terminan en muerte*.

Un informe interesante es el realizado por la Fundación Rock Health, "Big Data in digital health" [8], que revisa el estado actual y futuro del Big Data en el sector salud a través de una serie de entrevistas a empresarios, inversores y científicos. En este estudio se calcula que los **ahorros** en el sector sanitario debido a la adopción de Big Data pueden alcanzar de **325 a 525 millones de dólares al año**: 25-50 millones de dólares en la **mejora en la coordinación de la atención al ciudadano**; 125-175 millones en la **lucha contra los fraudes y los abusos**; 175-250 millones de dólares en la **lucha contra las ineficiencias administrativas y clínicas**.

Según el informe, hay seis vías mediante las cuales el Big Data puede cambiar la sanidad:

- Apoyo a la investigación (genómica y más allá).
- Transformación de datos en información.
- Apoyo al autocuidado de las personas.
- Apoyo a los proveedores de cuidados médicos.
- Aumento del conocimiento y concienciación del estado de salud.
- Agrupamiento de los datos para expandir el ecosistema.

Los principales retos a los que se enfrenta el sector para ello según este estudio son: la adopción cultural, el alineamiento de términos clínicos y técnicos, la sobrecarga de información, los silos de información, la inadecuada financiación, la falta de gobierno y el rechazo a la compartición de datos.

También señala tres tendencias en el uso de los datos que pueden tener cierta trascendencia: trabajar con conjuntos limitados de datos, mezclar estos conjuntos de datos limitados, combinar una gran variedad de datos, y agrupar los datos para obtener mejores resultados (*pooling*).

A futuro, el informe señala la existencia de máquinas que actuarán como **asistentes personales de los doctores** y que usarán Big Data para **ayudarles en la toma de decisiones**, pudiendo leer grandes cantidades de información de forma ultra-rápida, **monitorizando datos en tiempo real** (según van siendo publicados) e incluyendo **una gran variedad de fuentes de información de los pacientes**: registros electrónicos, genómica, datos clínicos, publicaciones, etc.

Para concluir este ámbito de aplicación, es importante destacar que uno de los principales obstáculos o preocupaciones existentes en la adopción de Big Data en salud tiene que ver con la **privacidad y la ética**. Aunque existen técnicas de anonimización de datos que eliminan los nombres y otra información personal de los registros sanitarios, se debe estar alerta y vigilar los posibles problemas que se puedan derivar en este sentido, sobre todo con la tendencia de apertura al público de la información. Así mismo, señalar que también es necesario un **cambio de mentalidad y una concienciación sobre el valor que aporta** el Big Data en este ámbito, tanto del lado de los pacientes como del de los especialistas médicos.

## TURISMO

Según apuntan todos los estudios, el sector turístico es uno de los ámbitos de aplicación donde mayor impacto tendrá la inversión en Big Data, cuyo foco está puesto en la **personalización de los servicios turísticos**, creando **experiencias enriquecedoras y adaptadas** para cada turista, y generando **nuevos modelos de negocio y servicios asociados**.

El recién publicado estudio "Big Data: retos y oportunidades para el turismo", de Invat.tur, el Instituto Valenciano de Tecnologías Turísticas, y Territorio Creativo [9], muestra cómo el concepto Big Data se introduce en este ámbito de aplicación unido directamente a la **gestión de la información y del conocimiento de las empresas y destinos**: la calidad y cantidad de datos que éstas puedan obtener para reforzar su toma de decisiones será el elemento

diferenciador para aumentar su competitividad. Además, destaca el nuevo contexto social y tecnológico, donde el **turista se sitúa en el centro del sistema** y es quien diseña, realiza y controla buena parte de la cadena de negocio gracias a Internet, esperando a cambio una relación menos comercial y un nivel de atención más humano, natural y personalizado. En este contexto, el Big Data permitirá **conocer al turista con gran precisión**, gracias al **análisis de datos de contexto en tiempo real**, permitiendo incluso adelantarse a sus decisiones y, por tanto, haciéndole sentir único y especial.

Algunos casos de éxito de aplicación de Big Data, tanto para la optimización de los procesos de negocio como para la mejora de la experiencia de usuario en alguna de las fases de su viaje, según dicho estudio, son:

- *Kayak: predicción a siete días de los precios de billetes de avión.* A diferencia de otros buscadores de viajes, incluye análisis predictivo en su módulo de vuelos para predecir si el precio del billete variará en los próximos siete días. Además, Kayak realiza numerosas pruebas Test A/B para optimizar su página web y la experiencia de usuario en la misma, donde se evalúan desde nuevos algoritmos de aprendizaje automático para la personalización de las tarifas hasta rediseños de la web o de la aplicación móvil.
- *Moovit: transformar la experiencia de usuario con open data.* Es una aplicación GPS de transporte público colaborativa y gratuita que informa sobre el estado del transporte público, permitiendo planificar rutas teniendo en cuenta los medios disponibles y facilitando la búsqueda de alternativas en situaciones de incidencia. Es una iniciativa crowdsourcing que combina datos generados por los usuarios con los datos que facilitan los operadores de transporte público (horarios, tiempos de llegada, etc.).
- *Landry's: detección de fraude de empleados.* La corporación Landry Inc. que agrupa empresas del sector de restauración, entretenimiento, juegos de azar y hostelería decidió atajar el problema del fraude de los empleados invirtiendo en tecnología Big Data para monitorizar en tiempo real los pedidos en restaurantes, consiguiendo un ahorro de 108 dólares semanales en cada restaurante así como un aumento de las ventas.
- *Museos: utilización del Big Data para conocer a los visitantes.* Existen numerosas iniciativas en Big Data cuyo objetivo es conocer cómo se comportan los visitantes de los museos y, en función de ello, decidir el diseño de las exposiciones, mejorar la publicidad y el marketing, o aumentar el número de visitas y de ventas en sus tiendas: museo de Arte de Dallas, el Instituto de Arte de Minneapolis, el Metropolitan de Nueva York, etc.

La encuesta sobre Big Data a profesionales del turismo en España realizada por Invat.tur refleja las siguientes conclusiones:

Se aprecia un **conocimiento más conceptual que práctico del Big Data** y una capacidad generalizada para apreciar sus beneficios y aplicaciones. La principal utilidad dentro del sector turístico identificada es la **predicción de la demanda**, seguida de un **mayor conocimiento sectorial** y de la obtención de **información en tiempo real**. En lo referente a las fuentes de datos que más interés suscitan son la **reputación online**, los intereses manifestados por los usuarios en los medios sociales y la información sociodemográfica. También se destacan las posibilidades de localización que ofrecen los dispositivos móviles (**geolocalización**).

En cuanto a las áreas donde Big Data puede generar un impacto más positivo en el sector turístico, los encuestados destacan como fundamental el hecho de **conocer con suficiente antelación los cambios en las preferencias y hábitos de consumo de la demanda**. Sin embargo, en general no han valorado significativamente los beneficios que puede suponer para la gestión de los propios servicios o para la optimización de los procesos internos.

Los requisitos que debe cumplir una empresa o destino turístico para maximizar las oportunidades que presenta el Big Data según dicha encuesta son el disponer de **profesionales cualificados** y acometer una correcta **planificación estratégica**.

En lo que respecta a la madurez tecnológica para la adopción de Big Data, se han indicado oportunidades de **evolución y mejora en la infraestructura y capacitación de las áreas de sistemas de información**, donde prevén en el corto plazo realizar las inversiones necesarias para dotarse de tecnologías novedosas e innovadoras.

La adopción de técnicas de análisis de grandes volúmenes de información entre las compañías entrevistadas del sector turístico ha alcanzado el 13%, destacando destinos turísticos y empresas de consultoría e investigación turística. **Más de la mitad de esta inversión se ha realizado en 2014**. Y entre las principales dificultades para ello se han destacado el **volumen de inversión requerida, la cultura organizativa y los profesionales adecuados**.

Por último, en cuanto al futuro, las empresas han señalado que Big Data facilitará la aparición o fortalecimiento de negocios basados en una **gestión predictiva de la demanda**.

Otro estudio interesante en este ámbito de aplicación es el realizado por Amadeus IT Group, "At the Big Data Crossroads: turning towards a smarter travel experience" [10] donde se indica de nuevo el **potencial del Big Data para transformar el negocio turístico y la experiencia de viaje**, ofreciendo beneficios como la **toma de mejores decisiones, productos y servicios nuevos, mejor relación con el cliente, y procesado de datos más rápido y barato**, lo que supone una mayor rapidez, eficiencia y eficacia en los procesos de negocio.

Los principales retos que indica el estudio para aprovechar el potencial del Big Data en el ámbito turístico son de tipo técnico y operativo: crear un repositorio de datos integrado, trabajar en un entorno tecnológico híbrido, superar la escasez de habilidades disponibles (poco personal cualificado en gestión y análisis de Big Data) y mantener la ventaja competitiva, mediante la creación de un plan de acción adecuado.

## FABRICACIÓN

Con la llegada de las **fábricas del futuro o la fabricación 4.0**, donde las tendencias tecnológicas predominantes giran en torno a la **colaboración, la movilidad, la conectividad y la inteligencia**, tal y como se recoge en el reciente informe "Las Tecnologías de la Información y las Comunicaciones para la Industria del Futuro" de ITI [11], los sectores manufactureros serán los **mayores productores de datos en tiempo real**.

Las fábricas del futuro posibilitarán una **fabricación sostenible medioambientalmente**, una **fabricación ágil y colaborativa**, una **fabricación centrada en las personas**, y una **fabricación óptima, innovadora y personalizada**. Se evolucionará de una producción simple y totalmente bajo control humano a un proceso complejo, automatizado y cada vez

más inteligente que **requiere de datos precisos y fiables**. Así, la adopción de estrategias para el análisis de grandes volúmenes de datos puede ser vital ya que, además de aumentar la **productividad de las empresas manufactureras**, el análisis de esta información permitirá **reducir costes, y predecir y prevenir problemas de funcionamiento**.

El uso de las TIC en la industria en general, y del Big Data en particular, será vital a la hora de alcanzar esta **sostenibilidad, excelencia operativa y eficiencia**, y se focalizará en las siguientes áreas según el citado informe:

- **Conexión del mundo físico en las plantas de producción.** Recursos reales como maquinaria, robots, líneas de producción, artículos y operadores forman parte de la estructura de información de los procesos de producción. Todos ellos deben estar conectados entre sí y con los sistemas finales empresariales, asegurando la transferencia de información y la concienciación del entorno que les rodea.
- **Minería y almacenamiento de datos.** Se deben almacenar sin fallos grandes cantidades de datos procedentes de la planta y de las cadenas de suministro; la información embebida en estos datos debe poder ser extraída y estar disponible.
- **Plataformas de servicios seguras, de alto rendimiento y abiertas.** Aplicaciones distribuidas y colaborativas serán implementadas mediante la mezcla de servicios de diferentes fabricantes y empresas TIC. La provisión de funcionalidades personalizadas a través de servicios se hará en la nube garantizando fiabilidad, seguridad y rendimiento.
- **Herramientas de modelado y simulación.** Los entornos complejos necesitan ser descritos con modelos semánticos para poder relacionar información, describir dinámicas y pronosticar comportamientos.
- **Arquitecturas de aplicación colaborativas y descentralizadas, y herramientas de desarrollo,** para soportar empresas y mercados globales donde las aplicaciones de gestión del ciclo de vida, de gestión de la cadena de suministro, de monitorización y de control, entre otras, no funcionan de forma estanca; y para reducir la complejidad y aumentar la flexibilidad entre los *stakeholders* en las cadenas de valor de fabricación, que requieren de información en tiempo real sobre los procesos, productos y cuellos de botella.

Según un artículo reciente de McKinsey<sup>9</sup>, los fabricantes pueden aprovecharse de la analítica avanzada para **reducir los fallos en los procesos productivos, a la vez que ahorran tiempo y dinero**. Esto es de especial interés en algunos entornos productivos como el farmacéutico, el químico o la minería, donde hay mucha variabilidad. La analítica avanzada se refiere a la aplicación de **herramientas estadísticas y matemáticas** sobre los datos para asesorar y mejorar las prácticas seguidas en los procesos de fabricación. Los gestores de operaciones pueden usar esta analítica avanzada para analizar profundamente el histórico de datos de los procesos, **identificar patrones y relaciones**, y para **optimizar los factores que tengan mayor efecto**.

Para aprovechar las ventajas que permite esta analítica avanzada, según McKinsey, el primer paso que deben dar los fabricantes es saber cuántos datos tiene la empresa a su disposición. Muchas empresas recogen grandes cantidades de datos de sus procesos pero sólo los usan

9

[http://www.mckinsey.com/insights/operations/how\\_big\\_data\\_can\\_improve\\_manufacturing](http://www.mckinsey.com/insights/operations/how_big_data_can_improve_manufacturing)

para monitorización y no como base para mejorar sus operaciones. Para estos, el reto es **invertir en sistemas y capacidades** que les permitan optimizar el uso de la información de sus procesos, por ejemplo, centralizando o indexando los datos de múltiples fuentes para que puedan analizarse más fácilmente, y contratando analistas de datos que estén formados especialmente en la detección de patrones y extracción de conclusiones.

Un ejemplo de fábrica inteligente en Europa que ya está aprovechando el Big Data es la que posee Siemens en Amberg (Baviera), una de las plantas más automatizadas del mundo. La compañía combina el análisis de datos con la inteligencia artificial. Además crea entornos tridimensionales mediante tecnologías 3D, de realidad virtual y sistemas ciberfísicos, para saber cómo mejorar la productividad. Gracias a esta gestión masiva de datos, la industria se simplifica y aumenta su competitividad. Además de aumentar su eficiencia energética, son capaces de adaptar mejor sus recursos a la demanda que prevén y llevar al mercado productos más complejos en mucho menos tiempo en un entorno cada más global y competitivo<sup>10</sup>.

Por último destacamos el estudio "Manufacturing and the Data Conundrum" de la Economist Intelligence Unit [12], que examina cómo los fabricantes recogen, analizan y usan los datos que se generan en tiempo real en los procesos de producción. Para ello se encuestó a 50 ejecutivos de la industria de Norteamérica y de Europa, incluyendo fabricantes de electrónica, maquinaria, productos químicos, componentes de vehículos y plásticos, entre otros.

Los aspectos más destacados de este estudio se listan a continuación:

- Se ha producido un **aumento significativo en la recolección de datos de las plantas de producción**, comparado con los dos últimos años.
- Una **minoría de fabricantes tiene ya una estrategia de gestión avanzada de datos**.
- Los fabricantes encuentran **dificultades a la hora de integrar datos de diversas fuentes y también para encontrar personal cualificado** para el análisis de los mismos.
- Mientras que la recolección de datos de monitorización es común, la **analítica de datos para predecir o solucionar problemas en la producción no lo es tanto**.
- Se han observado **ganancias** nada despreciables en cuanto **a la calidad y la eficiencia de producción**: dos tercios de las empresas entrevistadas han indicado ahorros anuales del 10% en términos de coste de calidad (pérdidas netas incurridas debido a los defectos) y en eficiencias de producción; y un tercio de las mismas ha indicado que sus ahorros en ambas medidas han estado entre el 11 y el 25%.

## BANCA Y FINANZAS

El sector financiero español está inmerso en un proceso de reestructuración y saneamiento sin precedentes motivado por la crisis económica de los últimos años. Son muchos los desafíos a los que bancos y cajas de ahorro tienen que hacer frente en la actualidad. Por un lado, la desconfianza de los clientes, que cada vez son más exigentes, tienen más capacidad para analizar los productos y servicios de la competencia a través de Internet y de las redes

<sup>10</sup> La fábrica más inteligente de Europa produce a base de 'big data' <http://www.elmundo.es/economia/2014/04/16/534d662c268e3efc2d8b457c.html>

sociales y, en general, son menos fiables y más sensibles al precio. Por el otro, los bancos están sometidos cada vez a más normas y regulaciones como Basilea III. Además, los depósitos escasean y las entidades pueden conceder menos créditos, por lo que las capacidades de este sector para generar negocio están limitadas actualmente. Para hacer frente a esta situación, las entidades financieras podrían aprovechar las ventajas que ofrece el Big Data, convirtiendo datos internos y externos en conocimiento útil, obteniéndose **mejoras de eficiencia en la provisión de servicios financieros y aumentando la satisfacción de los clientes**, en definitiva, aumentando su competitividad y su situación frente a nuevos competidores<sup>11</sup>.

En este sentido, son tres las áreas críticas que podrían mejorarse en el sector financiero gracias a Big Data<sup>12</sup>:

- *Visión 360º de los clientes.* Se debe **consolidar la información interna** que poseen los bancos sobre sus clientes (cuentas, productos y servicios contratados, operaciones en cajeros, banca online y móvil, compras con tarjetas...) **con la información externa** (por ejemplo, de redes sociales, webs o información pública procedente de plataformas open data), para **gestionarla de forma única** y así disponer de una **visión holística de los clientes**, que permita **recomendar y personalizar las ofertas comerciales**, mejorando la venta cruzada y teniendo a los clientes satisfechos y fidelizados. Además, gracias a Big Data las entidades podrían **conocer las causas de abandono de clientes**, lo que permitirá **optimizar sus mecanismos de retención** y focalizar sus esfuerzos en aquellos que resulten más valiosos.
- *Gestión de riesgos.* La optimización del riesgo en una entidad financiera significa una disminución de las provisiones necesarias en sus balances, lo que supone aumentar la confianza de los reguladores y sobre todo un aumento inmediato de los beneficios en sus cuentas de resultados. Entre los riesgos implícitos en este sector destacan: el riesgo de crédito, que podría reducirse aplicando Big Data al permitir un conocimiento integral de los clientes, **mejorando así los modelos de rating y scoring, y la gestión de la morosidad**; el riesgo de mercado, donde las soluciones Big Data permitirían **conocer en tiempo real lo que está aconteciendo en los mercados así como predecir lo que va a suceder en el futuro, mejorando la toma de decisiones**; el riesgo de liquidez, que podría verse minimizado con tecnologías Big Data que ayuden a **evaluar las entradas y salidas de fondos y faciliten la predicción de situaciones futuras**; el riesgo exterior, para el cual el Big Data posibilitaría un **diagnóstico claro de los riesgos derivados del comercio y la inversión en el exterior** y, por tanto, **predecirlos**; el riesgo operacional, donde el Big Data como gran recolector de la información que atañe a los procesos, sistemas internos y a la propia actividad de la entidad será clave para **analizar, predecir y ayudar a minimizar las pérdidas derivadas**; y el riesgo reputacional, que podrá ser medido, gestionado y predicho de forma más óptima.

<sup>11</sup> <http://www.aunclidelastic.com/big-data-como-solucion-a-los-desafios-del-sector-financiero/>

<sup>12</sup> <http://www.aunclidelastic.com/el-papel-de-big-data-en-tres-areas-criticas-del-sector-bancario/>

- *Gestión de fraudes.* Mediante Big Data las entidades financieras podrían implementar medidas para la **detección de actividades fraudulentas en tiempo real**, tanto internas como externas, destacando dos grupos: **detección de fraude en los medios de pago y prevención de blanqueo de capitales y financiación del terrorismo**. En el primer caso, mediante técnicas de Big Data las entidades podrán **detectar posibles patrones de comportamiento fraudulento** basándose en el análisis del histórico de datos en operaciones llevadas a cabo a través de cajeros, comercios, etc. En el segundo grupo, con Big Data los bancos podrán rastrear las operaciones sospechosas e incluso recurrir a fuentes externas para obtener una **información integral del individuo** que está llevando a cabo esos movimientos.

Destacamos el informe “El impacto del Big Data en los servicios financieros” de José García Montalvo, catedrático y director del departamento de economía y empresa de la Universitat Pompeu Fabra [13], donde se hace un repaso sobre las posibilidades que ofrecen las técnicas de Big Data en el campo de las entidades financieras, en línea con lo ya explicado. Entre ellas destaca la **microsegmentación**, la **calificación crediticia de los consumidores**, la **dinámica, predicción y recomendación de nuevos productos**, la **detección del fraude en tarjetas de crédito**, la **identificación de operaciones sospechosas** de blanqueo de capitales o actividades terroristas, la **gestión eficiente de las relaciones con los clientes** en un contexto de multicanalidad o la **fijación de objetivos de ahorro para los clientes** a partir del análisis de sus ingresos y gastos. Así mismo, se destaca que en el sector financiero son de especial relevancia en la gestión de los datos las cuestiones relacionadas con la **privacidad, la protección de datos y el cumplimiento de algunas regulaciones**. Se indica también que, como en cualquier proyecto empresarial, es importante analizar la relación coste-beneficio del proyecto Big Data así como contar con un equipo experto que sea capaz de extraer conocimiento a partir de los datos.

Por último, es interesante resaltar el convenio firmado el pasado 9 de marzo de 2015 por Daniel Peña, rector de la Universidad Carlos III de Madrid (UC3M), y Ana Botín, presidenta del Banco Santander, para la **creación del Instituto Mixto de Investigación en Big Data Financiero**, con sede en el campus de Madrid-Puerta de Toledo de la universidad. Este Instituto tiene como objetivo convertirse en un centro de liderazgo mundial en la recogida, almacenamiento, tratamiento y análisis de las grandes masas de datos que generan las instituciones, empresas y organizaciones gubernamentales. Trabaja con “Big Data” en todos los ámbitos con un foco especial en el sector de la banca y finanzas, a través de programas de formación, investigación y puesta en marcha de proyectos con empresas e instituciones públicas y privadas<sup>13</sup>.

13 [http://www.santander.com/csgs/Satellite/CFWCSancomQP01/es\\_ES/Corporativo/Sa-la-de-comunicacion/2015/03/09/La-UC3M-y-Banco-Santander-crean-un-Instituto-pionero-de-Investigacion-en-Big-Data-Financiero-.html](http://www.santander.com/csgs/Satellite/CFWCSancomQP01/es_ES/Corporativo/Sa-la-de-comunicacion/2015/03/09/La-UC3M-y-Banco-Santander-crean-un-Instituto-pionero-de-Investigacion-en-Big-Data-Financiero-.html)

## GESTIÓN URBANA/CIUDADES INTELIGENTES

Garantizar la seguridad ciudadana, gestionar el uso racional de los recursos, anticiparse a los problemas de movilidad urbana, y analizar la opinión y el sentir del ciudadano son solo algunos ejemplos de lo que el Big Data puede hacer en las ciudades del futuro<sup>14</sup>.

No podemos pasar por alto el cambio demográfico que se está produciendo en los últimos años, que reafirma ese papel de las nuevas tecnologías. Así, según el informe sobre el estado de las ciudades del mundo de ONU-HABITAT, se espera que **en 2050 el 86% de la población de países desarrollados viva en las ciudades**, siendo del 67% en las regiones menos desarrolladas. Globalmente se espera que 7 de cada 10 personas esté viviendo en un área urbana en 2050. Este panorama hace evidente la necesidad acuciante de mejorar la gestión y la eficiencia de las ciudades, por lo que las TIC en general, y el Big Data en particular, van a jugar un papel crucial.

En cuanto a la generación de datos, estamos viviendo una explosión debido a la abundancia de interacciones personales a través de redes sociales, terminales móviles, etc. y también a la existencia de millones de dispositivos M2M. En concreto, **las ciudades metropolitanas son los lugares con mayor concentración de personas y sensores**. Además, esta gran cantidad de información que se está generando en las ciudades es muy variada (vídeos, tuits, coordenadas GPS, archivos Excel, emails, etc.).

En definitiva, para que la ciudad sea realmente inteligente debe recolectar, procesar, compartir, almacenar y analizar una gran cantidad de datos que viene de distintas fuentes, para posteriormente transformarlos en conocimiento útil que será utilizado por los líderes de la ciudad para la toma de decisiones.

En este escenario, algunas de las áreas susceptibles de mejora gracias al uso del Big Data son las siguientes:

- **Seguridad ciudadana.** Se podría mejorar la eficiencia de los servicios de policía y bomberos mediante la correlación de los datos que proceden de todos los sistemas instalados en la ciudad, incluyendo cámaras de vigilancia, sensores de humo y de fuego, y la geolocalización de vehículos de emergencia. Las tecnologías de Big Data podrían usarse en el ámbito de la prevención de delitos en la ciudad para coordinar mejor las respuestas, para la provisión eficiente de servicios sociales y para una mejor comprensión del panorama criminal.
- **Transporte urbano.** Mediante la captura de datos en tiempo real de las cámaras de vídeo y los sensores magnéticos instalados en la red de carreteras, podríamos monitorizar la localización de los autobuses públicos y utilizando analítica predictiva se podría comprender mejor y predecir patrones humanos y de tráfico, lo que ayudaría a disminuir las interrupciones del servicio y a una mejor gestión de los sistemas de transporte. De la misma forma, monitorizando las redes sociales se podrían detectar futuras manifestaciones y,

14 <http://www.aunclidelastic.com/big-data-piedra-angular-de-las-smart-cities/>; <http://blog.digital.telefonica.com/blog/2014/05/21/big-data-smart-cities/>

en consecuencia, facilitar la gestión de los potenciales atascos cambiando las rutas de los autobuses, modificando las secuencias de los semáforos e incluso transmitiendo información a los teléfonos móviles de los conductores indicando los tiempos de conducción previstos así como rutas alternativas.

- **Gestión del agua.** A través del análisis de los datos que proceden de los sistemas de medición como los sensores de presión, de PH y de turbidez del agua instalados en los sistemas de abastecimiento y saneamiento así como de las videocámaras situadas en las plantas de tratamiento de agua, sería posible optimizar la gestión del agua mediante la detección de fugas, la reducción del consumo de agua y la minimización de la inundación de las alcantarillas por ejemplo.
- **Energía.** Es posible optimizar la producción de la energía, adaptándola a la demanda en tiempo real, mediante la recolección de los datos que proceden de los contadores eléctricos inteligentes instalados en las viviendas así como los datos que proceden de plataformas open data y las previsiones meteorológicas. Esto supone un importante ahorro energético y un crecimiento más sostenible y ecológico. En concreto, según la Economist Intelligence Unit, tres de cada cuatro eléctricas europeas esperan que el volumen de datos crezca un 25% en los próximos tres años, pero la mitad de ellas afirman que actualmente no están maximizando los datos que están recogiendo. Entre las posibilidades que ofrece el Big Data para estas empresas destacan: conocer de primera mano cuándo se dan los picos de demanda, de tal forma que aumenta su eficiencia; mejorar el tiempo de respuesta ante una interrupción del servicio eléctrico; y priorizar las operaciones de mantenimiento así como conocer situaciones de fraude o robo. Pero sobre todo, el mayor beneficiario puede ser el cliente, permitiendo a las eléctricas personalizar su experiencia y ofrecer paquetes de servicios *ad hoc* que satisfagan las necesidades de cada cliente<sup>15</sup>.
- **Residuos urbanos.** Si utilizamos los datos en tiempo real que proceden de los sensores que detectan el llenado de los contenedores y lo comparamos con el histórico de datos, podríamos predecir el periodo de vaciado de los contenedores así como la optimización de las rutas de recogida de basura.
- **Análisis del sentir del ciudadano.** A través del análisis de las redes sociales y blogs mediante tecnologías Big Data, se podría medir la opinión pública de ciudadanos y turistas en aspectos y servicios claves de la ciudad como el transporte público, la gestión de residuos o la seguridad ciudadana, lo que ayudaría a priorizar y a adecuar las políticas municipales derivadas.
- **Transparencia y Buen Gobierno.** Uno de los principales pilares del *open government* o gobierno abierto es la transparencia, es decir, la obligación de las administraciones de dar cuenta a los ciudadanos de todas sus acciones, con especial atención de aquellas que requieran del uso de dinero público, y prevenir así los casos de corrupción y mal gobierno. La vía directa para que un gobierno sea abierto es poner a disposición de los ciudadanos los datos que generan las administraciones públicas (open data) de un modo legible, asequible, comprensible, estructurado y reutilizable. El volumen exponencial de datos abiertos hace necesario el uso de sistemas Big Data que puedan capturar y procesar esos volúmenes de información puesta a disposición de la ciudadanía, y por extensión al propio gobierno, y permita a los responsables públicos mejorar sustancialmente la toma de decisiones y la adopción de distintos tipos de medidas económicas, sociales, etc. Las iniciativas de gobierno abierto necesitan del uso

de las TIC, y en especial de Big Data, para llegar a un nivel de transparencia real y efectiva acorde con la sociedad del siglo XXI y el deseo de la ciudadanía<sup>16</sup>.

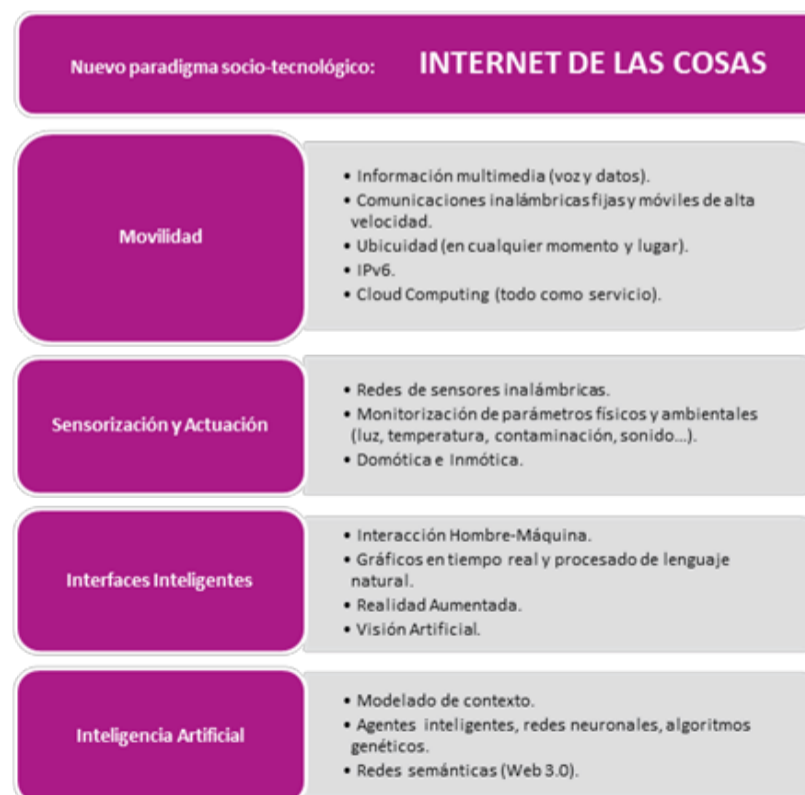
- **Educación.** Utilizando minería de datos y analítica avanzada en educación se podría aumentar considerablemente la calidad de la enseñanza, la monitorización y la evaluación. Por ejemplo, se podrían identificar las deficiencias existentes en los colegios en cuanto al rendimiento del alumnado, a su seguridad y al grado de preparación para el mundo laboral, adaptando y personalizando las soluciones para mejorar los resultados educativos.

En todos estos entornos de aplicación es fundamental el **Internet de las Cosas**. Se deben instalar cientos de sensores en edificios públicos (de seguridad, de alumbrado), en sistemas de gestión de energía (contadores, turbinas, generadores, baterías), en plataformas de transporte (vehículos, semáforos, señales) y en los sistemas de seguridad (ambulancias, coches de bomberos...).

Todos estos datos se transmitirán a un servidor central que los correlacionará y analizará combinados con otras fuentes de datos para al final obtener información útil. Por ello, también será necesario considerar las tecnologías de **Cloud Computing**, que permiten reducir costes, un menor tiempo de provisión al mercado y, por supuesto, mayor flexibilidad y escalabilidad.

En definitiva, **el término Ciudad Inteligente va necesariamente ligado a los términos de Big Data e Internet de las Cosas.**

En el informe "Smart Environments: Las TIC en las Ciudades Inteligentes" de ITI [14] se profundiza en este conjunto de tecnologías facilitadoras de la ciudad inteligente en cuatro dominios: *movilidad, sensorización y actuación, interfaces inteligentes e inteligencia artificial*, tal y como se muestra en la siguiente figura resumen:



*Figura 20. Tecnologías facilitadoras de la Ciudad Inteligente (Fuente: ITI [14])*

Son numerosas las ciudades que están adoptando iniciativas de Big Data, destacando entre ellas **Chicago** (Estados Unidos), que está **a la vanguardia de aplicar la gestión de datos**

16 Gobierno abierto y transparencia de la mano del "big data" [http://www.astic.es/sites/default/files/articulosboletic/mono9\\_laura\\_preciado.pdf](http://www.astic.es/sites/default/files/articulosboletic/mono9_laura_preciado.pdf)

**masivos en transporte, salud, educación, energía, seguridad ciudadana y otros retos urbanos.** Algunos de sus proyectos destacados son los siguientes<sup>17</sup>:

- *Chicago Health Atlas*, donde los ciudadanos pueden visualizar en mapas las tendencias de salud en la ciudad de Chicago así como visualizar estadísticas específicas y encontrar recursos sanitarios en sus barrios.
- *Chicago Energy Data Map*, que permite informar a los ciudadanos del consumo de electricidad y gas natural de forma visual y sencilla. El mapa compara los ratios de eficiencia energética por áreas o zonas residenciales a lo largo de los distintos barrios de Chicago. Además se muestran trucos rápidos para ahorrar energía.
- *Second City Zoning*, donde se muestra de forma sencilla información sobre lo que se puede construir y dónde. Este mapa permite a los usuarios detectar patrones en cuanto al desarrollo urbanístico e incluso al detalle de barrios.
- *SchoolCuts*, que fue creado como repositorio de información sobre los colegios públicos de Chicago. El sitio presenta datos demográficos, de rendimiento y de inscripción en los diferentes colegios de la ciudad.

El reciente estudio "The Emerging Power of Big Data: The Chicago Experience", del Chicago Council on Global Affairs [15], pone de manifiesto cómo utilizar Big Data para la toma de decisiones municipales y de gobierno mediante un caso de uso real, la ciudad de Chicago, catalogada como una de las mega ciudades del mundo por prever una población de más de 10 millones en 2025. En el estudio se revisan los que consideran los cuatro pilares de la gestión urbana: la energía, el transporte, la educación y la seguridad ciudadana.

Como ejemplo de solución en este ámbito de aplicación, destacamos la *aplicación móvil Bigov*, de la empresa catalana Bismart, que ha recibido el premio Microsoft a la app más útil del mundo para los gobiernos a través de su aplicación móvil *Bigov Better City Indicators*. Esta aplicación transforma los datos que se generan en las ciudades en información de valor que ayuda a los organismos públicos en la toma de decisiones y a crear ciudades más habitables y sostenibles. La aplicación ofrece una visión en tiempo real de 80 indicadores clave de rendimiento de las ciudades, desde la economía local, pasando por el nivel de delincuencia hasta la educación<sup>18</sup>.

## MEDIOS DE COMUNICACIÓN Y ENTRETENIMIENTO

La crisis económica, unida a los efectos de Internet en el modelo de negocio tradicional, ha golpeado de lleno en el sector de los medios de comunicación. La caída de la inversión publicitaria, la entrada de nuevos competidores en el mercado y los cambios en los patrones de consumo han puesto en jaque a un sector que desde 2008 ha visto como, solo en España, han cerrado cerca de 300 medios de comunicación, siendo los principales afectados los medios en papel (revistas, periódicos, etc.) y las televisiones<sup>19</sup>.

Frente a este panorama, ¿pueden las tecnologías Big Data cambiar un sector como el de los medios de comunicación? ¿puede este sector beneficiarse de la revolución de los datos? La

17 Chicago: City of Big Data <http://bigdata.architecture.org/what-is-big-data/>

18 Una 'app' española de gestión urbana, la mejor del mundo <http://www.elcorreo.com/alava/tecnologia/empresas/201410/15/espanola-gestion-ciudad-mejor-20141015150008-rc.html>

19 [http://sociedad.elpais.com/sociedad/2013/12/12/actualidad/1386862599\\_529386.html](http://sociedad.elpais.com/sociedad/2013/12/12/actualidad/1386862599_529386.html)

respuesta es sí, y todos los esfuerzos e inversiones en tecnología por parte de las empresas irán dirigidos a un aspecto primordial en su negocio: **conocer más y mejor a su usuario o cliente**, a través de la **recolección y análisis de datos que posee la propia empresa (audiencias)**, así como aquellos que **generan los propios clientes en diferentes canales como las redes sociales**. El objetivo es permitir que los medios puedan trabajar en **contenidos cada vez más personalizados**, dando a sus lectores o espectadores aquellos que quieren o buscan, y **mejorando la relación con los anunciantes**, al ser más eficaces a la hora de segmentar a su audiencia gracias al tratamiento de los datos generados.

La tendencia en el sector es clara. Según un estudio de EY, "Sustaining digital leadership" [16], el 70% de los directivos de la industria de los medios y el entretenimiento está dispuesto a realizar inversiones, incluso asumiendo pérdidas, para convertirse en referentes en el uso de nuevas tecnologías y ser líderes en el ámbito digital. En el estudio se identifica al Big Data como una de las herramientas más importantes para la transformación digital, cuyo uso se encuentra totalmente en fase primaria de adopción en el sector.

No obstante, las principales empresas líderes en el ámbito de los mass media ya están empezando a explotar y beneficiarse de la captura, tratamiento y análisis de grandes cantidades de datos. Como ejemplos podemos citar los siguientes:

- El Huffington Post utiliza tecnologías Big Data para optimizar sus contenidos. Los redactores cuentan con un dashboard que muestra tendencias de consumo en tiempo real y que permite adaptar los contenidos a su audiencia y mejorar la programación para aumentar el tráfico<sup>20</sup>.
- La estadounidense BuzzFeed contrató en 2010 a su primer científico de datos para predecir cómo y cuándo un contenido se convertiría viral en Internet. A día de hoy, la empresa cuenta con un equipo de más de 100 "data scientist" y ha invertido en una infraestructura tecnológica que permite soportar las necesidades de estos profesionales en materia de recolección y análisis de grandes cantidades de datos<sup>21</sup>.
- Bloomberg, medio especializado en información económica y financiera, ha puesto en marcha Bloomberg Denizen, una solución dirigida a sus anunciantes que pone en valor el conocimiento de la empresa en la explotación de los datos de su audiencia. El equipo de trabajo asociado está compuesto por periodistas y analistas de datos (data scientist) para ofrecer contenido segmentado a clientes potenciales de los servicios y productos de los anunciantes<sup>22</sup>, intentado rentabilizar al máximo la inversión en publicidad.
- Netflix, empresa estadounidense de contenido bajo demanda en Internet, tiene, entre las claves de su éxito, el uso de tecnologías Big Data. Siendo una de las pioneras del sector, Netflix apostó por algoritmos de análisis de los datos que generaba su audiencia, lo que le permitió conocer gustos, preferencias, patrones, etc. respecto a géneros, actores, incluso colores de portadas, según diferentes parámetros de sus clientes, y adaptar su oferta y realizar recomendaciones personalizadas. El análisis de estos datos, junto al cruce de otras fuentes, permite a la empresa detectar comportamientos futuros de su audiencia y optimizar las

20 <https://www.journalism.co.uk/news/how-the-huffington-post-uses-data-in-the-newsroom-and-beyond/s2/a556679/>

21 <http://www.buzzfeed.com/daozers/how-buzzfeed-thinks-about-data-science#.dqegaP7KL>

22 <http://www.bloomberg.com/company/press-releases/bloomberg-denizen-new-data-driven-content-solution-advertisers/>

inversiones en la producción de series de ficción<sup>23</sup>.

Por último, es interesante resaltar que las tecnologías Big Data están revolucionando la manera en que los periodistas realizan su trabajo. El llamado “**periodismo de datos**” puede definirse como una nueva disciplina de los medios de comunicación que ofrece a los lectores representaciones gráficas para explicar las noticias de forma clara y comprensible. Las tecnologías y, sobre todo, las herramientas de Big Data combinadas con fuentes de Open Data, están permitiendo a los periodistas ofrecer nuevos contenidos e historias de valor que se esconden tras los datos y que es necesario comunicar al público.

## 4.2 Barreras de entrada

Según el ya citado informe de IBM, existe un reto sistemático que obstaculiza la adopción de Big Data, y que es independiente de la fase de adopción en la que se encuentre la empresa: la capacidad para **articular un caso de negocio convincente**, tal y como muestra la siguiente figura. En el contexto socio-económico actual ya no es aceptable una inversión en tecnología si no se cuantifica su beneficio, requisito no exclusivo de las iniciativas de Big Data. Una vez conseguido ese primer punto, el siguiente desafío según IBM es encontrar las **habilidades necesarias** para que el Big Data resulte operativo, incluyendo habilidades **técnicas, analíticas y de gobierno**.

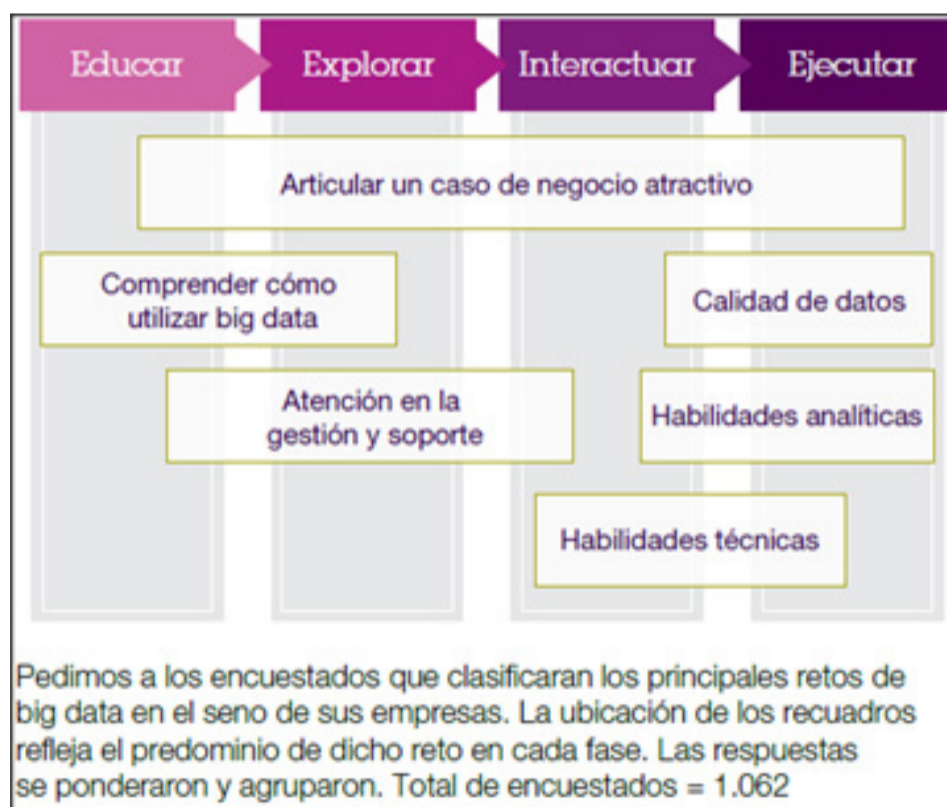


Figura 21. Obstáculos clave para la adopción de Big data (Fuente: IBM [1])

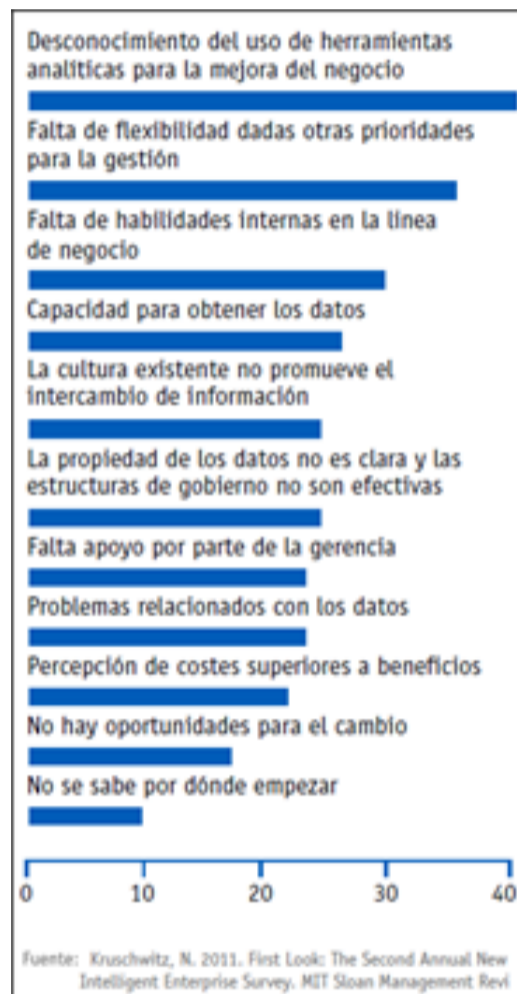
Por su parte, el estudio del OBS indica como retos a superar por las empresas que tienen planeado afrontar proyectos de Big Data los siguientes:

- **obtener verdadero valor** del Big Data,

23 <http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?pagewanted=all&r=0>

- definir una **estrategia para la recopilación del conocimiento**,
- **gestionar el talento, las habilidades y capacidades necesarias**,
- recopilar o definir **planes de inversión y financiación**, además de **cuestiones de liderazgo y organización**.

En la siguiente figura se muestran las principales barreras que se encuentran las empresas en el momento de planificar o abordar proyectos de Big Data acorde al estudio del OBS:



*Figura 22. Barreras en la adopción de Big Data (Fuente: OBS)*

Destacamos también el 5º informe anual de la consultora PwC, "Digital IQ Survey", donde se entrevistó a 1.108 personas de 12 países y de una variedad de industrias, tanto profesionales IT como líderes empresariales, y donde se encontró que el 62% de los encuestados creía que los grandes volúmenes de datos podían darles una ventaja competitiva, a la vez que el 58% estaban de acuerdo en que la compresión de datos era un gran desafío. Además, se descubrió que las organizaciones luchan con **cuatro grandes barreras del Big Data**<sup>24</sup>:

- Son **ciegas a la importancia de la visualización**. Para entender la realidad derivada de los datos, la visualización es fundamental, ya que ayuda a poner los datos en contexto y a hacer que los casos de negocio tengan sentido, llegando incluso a recopilar ideas que de otra forma no sería posible. Según el director de PwC el tipo de visualización hacia el que la mayoría se está moviendo consiste en tener paneles de información donde se puede acercar o alejar la imagen, pero este análisis es básicamente retrospectivo. Actualmente lo que se necesita es una visualización mucho más dinámica, que mire hacia el futuro.

- Se **interesan más en la recopilación de datos que en su análisis**. Las empresas están invirtiendo bastante en la recopilación de datos, pero no tanto en integrarlos, combinarlos y analizarlos: según el estudio, el 32% han invertido más de un millón de dólares en la recolección, almacenamiento y recuperación de datos internos, pero sólo el 26% han invertido esa cifra en su análisis. Se indica que los servicios financieros, las industrias de la salud y los seguros son los que parecen estar invirtiendo más en la integración de datos. En este contexto, se indicaba que debe haber un nuevo enfoque organizacional, que significa nuevos talentos y nuevas maneras de explotar los datos.
- Están enfrentando una **brecha de talento**. Las empresas a menudo carecen de talento en las habilidades necesarias para interpretar grandes volúmenes de datos: sólo el 44% de los encuestados de PwC dijo tener una cartera de talento suficiente como para llevar a cabo un análisis profundo de los datos. Se indica cómo las personas del análisis de marketing, los grupos actuariales y los desarrolladores de productos/precio pueden servir como buen punto de partida para convertir los datos en conocimiento.
- Están luchando con **sistemas insuficientes para procesar rápidamente la información**. La cuarta barrera de Big Data son los sistemas existentes, dado que los grandes volúmenes de datos exigen un mayor poder de cómputo para recopilar, almacenar y analizar los datos rápidamente. Muchas organizaciones dudan de su capacidad para hacerlo con los sistemas actuales: el 41% de los encuestados en las Américas, el 33% de los encuestados en Europa y el 49% de los encuestados de Asia-Pacífico.

Por último, la Big Data Value Association también ha analizado los aspectos a los que Europa debe enfrentarse para crear un ecosistema fuerte de Big Data, que abarcan las siguientes dimensiones:

- **Datos**. La **disponibilidad y el acceso a los datos es primordial**. Existe una gran variedad de tipos y fuentes de datos. El valor se genera mediante la adquisición, combinación y provisión de acceso a los datos de forma no retardada y asegurando la integridad y la privacidad de los mismos. Por último, se añade valor mediante el pre-procesado, la validación, el análisis y asegurando la integridad y precisión de los mismos.
- **Habilidades**. Un reto clave para sacar potencial al Big Data en Europa es asegurar la disponibilidad de **personal altamente y adecuadamente capacitado en metodologías y tecnologías** que permitan desplegar las aplicaciones y soluciones de Big Data. Se necesitarán científicos e ingenieros de datos que tengan experiencia en analítica, estadística, *Machine Learning*, minería de datos, y gestión de datos. Estos expertos técnicos deberán combinarse con expertos de dominio que tengan un fuerte conocimiento industrial y la habilidad de aplicarlo dentro de la organización para crear valor.
- **Legal**. Aspectos como la **propiedad de los datos y su uso, la protección y privacidad, la seguridad, la responsabilidad, el cibercrimen, los derechos de propiedad intelectual (IPR-Intelligence Property Rights)** y el impacto de bancarrotas en los derechos de los datos deben ser resueltos para eliminar barreras de adopción y posibilitar entornos regulatorios favorables.
- **Técnico**. Para desarrollar nuevas oportunidades y ventajas competitivas deben afrontarse algunos **avances científico-tecnológicos**, tal y como fue detallado en el apartado anterior: **analítica en tiempo real, procesamiento de datos escalable y con baja latencia, interfaces de usuario nuevas y enriquecidas, interacción de datos y enlace de datos, información y contenido**. Además la **interoperabilidad** de los conjuntos de datos y de las soluciones basadas en datos es fundamental para una amplia adopción. Por último, los **estándares de facto** también son un mecanismo primordial para evitar largos procesos de negociación que pueden desacelerar la interoperabilidad del mercado del Big Data.

- *Aplicación.* El negocio y las aplicaciones listas para el mercado deben ser un objetivo principal si se desea tener suficiente impacto y garantizar el liderazgo europeo en la creación de valor en Big Data.
- *Negocio.* Se deben desarrollar **ecosistemas y modelos de negocio apropiados** para esos ecosistemas si se desea tener un impacto positivo en el empleo y la economía Europea.
- *Social.* Es crítico para acelerar la adopción del Big Data **aumentar la concienciación sobre los beneficios y el valor que el Big Data puede crear** en el sector empresarial, el sector público y para los ciudadanos.

## VISIÓN Y CONTRIBUCIÓN DE ITI AL BIG DATA

### 5.1 Descripción general de ITI

El Instituto Tecnológico de Informática, ITI, es un **Centro Tecnológico especializado en Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones**. Desarrolla su actividad desde su creación en 1994 con clara orientación a la generación de nuevas tecnologías y metodologías para su transferencia a la industria y a la sociedad.

Como centro tecnológico de referencia, forma parte de las principales estructuras a nivel regional y nacional relacionadas con la generación y transferencia de tecnologías en general, destacando en este sentido su pertenencia al Instituto Valenciano de Competitividad Empresarial (IVACE), a la Red de Institutos Tecnológicos de la Comunitat Valenciana (REDIT) o a la Federación Española de Centros de Investigación y Tecnología (FEDIT), así como a estructuras relacionadas con la definición de las estrategias de investigación en Tecnologías de Investigación y las Comunicaciones, como las plataformas tecnológicas españolas y europeas.

De esta forma, su misión es **Investigar, Desarrollar e Innovar en Tecnologías de la Información y de las Comunicaciones para mejorar y mantener la posición competitiva de las empresas tecnológicas, generando y transfiriendo los conocimientos necesarios para la evolución de la industria y de la sociedad en general**.

La visión de futuro del Instituto es ser líder a nivel internacional en soluciones optimizadas e inteligentes basadas en Tecnologías de la Información y las Comunicaciones.

Esto se hace posible, gracias al cuidado proceso de I+D+I, que parte de la vigilancia tecnológica y de mercado para la detección de necesidades y oportunidades, el desarrollo de la I+D+I para generar conocimientos, tecnologías y metodologías y la conversión de éstas en soluciones que puedan llegar al mercado.

Actualmente el Instituto cuenta con una red de asociados formada por las principales empresas involucradas en la realización de actividades de I+D+I en torno a las TIC. Dentro del cuerpo de empresas asociadas se encuentran principalmente empresas regionales, que se basan en la cercanía geográfica para aprovecharse de los servicios que el Instituto presta, y además, empresas nacionales de referencia con las que el Instituto colabora de forma continuada.

ITI desarrolla su actividad en proyectos de ámbito regional, nacional e internacional, contando con colaboradores a todos estos niveles. Es muy representativa su presencia en las plataformas tecnológicas, tanto a nivel nacional como europeo, lo que le permite detectar las demandas del mercado e intervenir en la definición de las futuras líneas de investigación. Esta apuesta estratégica supone una importante ventaja competitiva para el Instituto en particular y para la industria TIC de la Comunidad Valenciana en general.

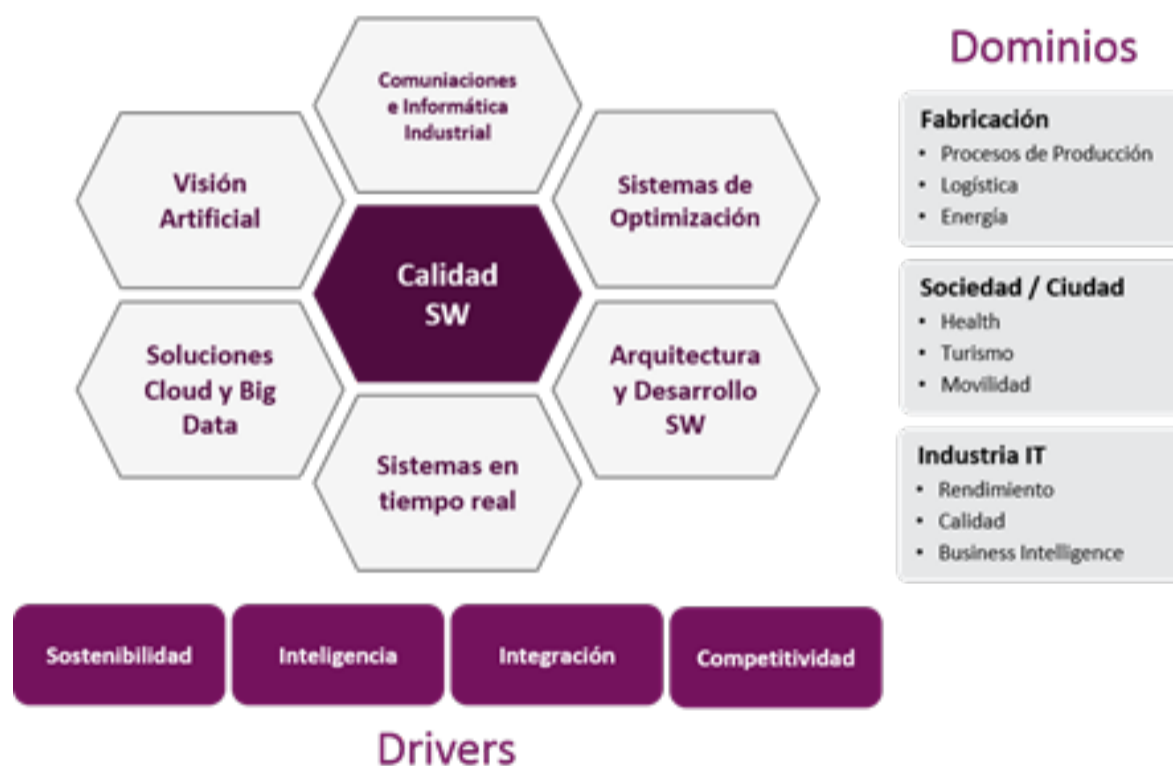
Entre todas ellas, destaca su contribución en PLANETIC, Plataforma Tecnológica Española para la adopción y difusión de las tecnologías electrónicas, de la información y la comunicación, que nace como la fusión de las plataformas INES, PROMETEO y GENESIS. ITI es miembro del comité de dirección de dicha plataforma y participa en los grupos de trabajo de Software y Servicios y de Sistemas Embebidos.

Además de PLANETIC, ITI está participando en muchas otras Plataformas Tecnológicas Españolas como eSEC, de Tecnologías para la Seguridad y la Confianza; es.INTERNET, de Convergencia hacia Internet del Futuro; Logistop, centrada en la Logística Integral; la Plataforma Tecnológica del Agua y del Riego, para la gestión sostenible de los recursos hídricos, la Plataforma Tecnológica Marítima, por una industria marítima sostenible y saludable; enerTIC, plataforma de TIC para la mejora de la eficiencia energética; eNEM, de Tecnologías Audiovisuales en Red; eVIA, para la Vida Independiente y Accesible, y ManuKET, de Fabricación Avanzada.

Así mismo, el Instituto está colaborando en algunas plataformas tecnológicas europeas que como son NESSI, ARTEMIS (ahora ECSEL) y Net!Works (ahora NetWorld2020).

Para desarrollar su actividad, ITI cuenta con una estructura organizativa compuesta por más de 80 profesionales, en su mayoría investigadores y tecnólogos, con un 87% de titulados superiores y un 14% de doctores.

ITI desarrolla su actividad de I+D+I enmarcada en las siguientes áreas, totalmente alineadas con el programa Horizonte 2020, la Estrategia Española de Ciencia, Tecnología y de la Innovación, la Agenda Digital Europea, Española y de la Comunidad Valenciana y la Estrategia de Especialización Inteligente en investigación e Innovación de la Comunidad Valenciana (RIS3 CV):



*Figura 23. Áreas de especialización TIC de ITI*

Con el centro de gravedad en la Calidad del Software, y con el foco puesto en los drivers que guían la investigación (sostenibilidad, inteligencia, integración y fomento de la competitividad), ITI desarrolla tecnologías y algoritmos de optimización (**Forecasting, Scheduling y Routing**) y Visión Artificial (**Inspección industrial en 3D e Inspección superficial por visión**). Así mismo, el Instituto cuenta con amplia experiencia en tecnologías que permiten la captación, comunicación y procesado de información a través de Internet, destacando

las capacidades de diseño de arquitecturas **Cloud Computing para el tratamiento de grandes cantidades de datos (Big Data)**, de diseño de Redes de comunicaciones (**redes de sensores, comunicaciones industriales, etc.**) y los *Cyber-physical systems* (**software empotrado, sistemas en tiempo real, verificación y validación de sistemas críticos, etc.**).

ITI es un centro muy horizontal en lo que a sus investigaciones respecta, pero se identifican una serie de sectores o dominios de aplicación en los que ITI trabaja habitualmente:

- Fabricación inteligente: Aplicación de soluciones avanzadas al entorno de fabricación para mejora de procesos, logística y eficiencia energética.
- Sociedad Inteligente: Aplicación de tecnologías avanzadas a la sociedad/ciudad, en ámbitos como salud, turismo o movilidad.
- Industria IT: Aplicación de tecnologías para mejora del rendimiento, la calidad o la inteligencia de negocios.

## 5.2 Capacidades y experiencias relacionadas

En el contexto del Big Data, ITI es miembro fundador de la **Big Data Value Association** ([www.bdva.eu](http://www.bdva.eu)) y centra su actividad en Big Data en tres ámbitos principales:

- Plataformas e Infraestructuras que soportan el manejo de grandes cantidades de datos
- Análisis de la información (*Data Science*)
- Visualización orientada a Estrategia/Valor (*Big Data Visualization & Insights*)

### Plataformas e Infraestructuras que soportan el manejo de grandes cantidades de datos

En esta línea, ITI trabaja intensamente en el diseño y despliegue de sistemas y bases de datos escalables, consistentes y fiables para permitir la construcción de aplicaciones distribuidas que cumplan unos niveles de calidad de servicio fijados. Cobra especial importancia en esta línea el modelo **Cloud Computing**, en el que ITI se centra en la investigación y diseño de patrones arquitectónicos y sus características de escalabilidad para definir modelos de servicio que faciliten el desarrollo de aplicaciones SaaS y automaticen su despliegue y gestión sobre las infraestructuras (IaaS) existentes, obteniendo el rendimiento necesario de las mismas para mantener los Acuerdos de Calidad de Servicio (SLA) establecidos.

Además, en este ámbito se está trabajando en *Ingeniería de Datos Almacenados*, incluyendo la mejora de la calidad, la integridad y la consistencia de datos, y el rendimiento de bases de datos. Existen distintas formas de **consistencia**: semántica (integridad), transaccional (ACID - Atomicidad, Consistencia, Aislamiento y Durabilidad), distribuida, de replicación, eventual (muy importante en Big Data) y parcial. Por su parte, la **calidad** de datos incluye cuantificar y comparar el tamaño de inconsistencias o consistencias en distintos conjuntos de

datos o sistemas (con restricciones o criterios de calidad).

Así mismo, se está trabajando en:

- **NoSQL** (“not only SQL”): sistemas de gestión de bases de datos que difieren del modelo clásico del sistema de gestión de Bases de Datos Relacionales (RDBMS) en aspectos importantes, entre ellos, el que no usan SQL como el principal lenguaje de consultas. Además, los datos almacenados no requieren estructuras de tablas fijas, normalmente no soportan operaciones JOIN, ni garantizan completamente ACID, pero a cambio de todo esto aportan capacidades de escalado mucho mayores y se adaptan mejor a fuentes de información desestructuradas o con una estructura no conocida a priori.
- **NewSQL**: se trata del nuevo paradigma bajo el que se agrupan los nuevos modelos de gestión de Bases de Datos Relacionales altamente escalables. Mientras que NoSQL mantiene unas filosofías de almacén y acceso a la información alternativas, NewSQL explora las posibilidades de uso de esquemas relacionales para grandes volúmenes de datos. Su aplicación es directa en caso de contar con datos bien estructurados pero cuyo volumen o tipo de uso hace necesario un almacén escalable, en algunos casos distribuido geográficamente (posiblemente en la nube), e incluso con necesidades de multi-arrendatario. NewSQL incorpora tecnologías que simplifican muchísimo todas estas gestiones, permitiendo de forma sencilla conseguir altos rendimientos incluso para soluciones que requieran *in-memory* para procesamiento en tiempo real.

## Análisis de la información (Data Science)

ITI tiene una amplia experiencia en la gestión y análisis de datos, sobre todo aplicando técnicas de reconocimiento de patrones y aprendizaje (*Machine Learning*) en grandes repositorios de información. Esta experiencia, sumada a un profundo conocimiento en las últimas tecnologías e infraestructuras Big Data, ha permitido a ITI agregar el prefijo **Fast Scalable** a las técnicas de *Machine Learning* que se manejaban hasta ahora para generar conocimiento/valor a partir de los datos.

Las más extendidas técnicas de análisis (como GLM –Generalization of Linear Regression o Redes de Aprendizaje Profundo –Deep Neural Networks, por citar alguna de las más representativas), ampliamente explotadas por ITI en distintos campos de aplicación, se benefician ahora de la facilidad para aportar distribución y paralelismo que aportan las tecnologías Big Data.

Los grandes volúmenes de datos hacen mejores a los algoritmos. Así, ITI combina la potencia de algoritmos avanzados con la capacidad de procesamiento *in-memory* realmente escalable para Big Data, en uno o en múltiples nodos de procesamiento. El resultado de todo esto es la creación de software basado en modelos predictivos para la detección rápida de anomalías, identificación de tendencias, o incluso para la generación de predicciones basadas en la información ya disponible en la empresa.

En definitiva, potentes algoritmos o técnicas como *Vector Support Machines*, regresión lineal general, análisis de series temporales, métodos de los K Vecinos más cercanos, análisis de componentes principales, análisis factorial, análisis de correspondencia, regresión de mínimos cuadrados parciales, algoritmos genéticos, programación lineal, el análisis multi-variante, control de calidad estadístico, fiabilidad estadística, regresión logarítmica lineal, análisis de

agregación, etc., explotadas y aplicadas por ITI a distintos ámbitos, se benefician ahora de la potencia de procesamiento necesaria para analizar grandes cantidades de datos, más modelos y más variables. En definitiva, Big Data permite a ITI hacer todo esto más rápido, más fácilmente, y de forma más rentable, maximizando el beneficio final para la empresa.

Destacar también que en esta línea se trabaja en la recolección de información de múltiples fuentes heterogéneas y la indexación inteligente, usando técnicas de clasificación semántica.

## **Visualización orientada a Estrategia/Valor (Big Data Visualization & Insights)**

La experiencia de ITI en el análisis de datos y en el desarrollo de interfaces centradas en la experiencia de usuario, le permiten abordar de forma eficaz el reto de desarrollar sistemas de apoyo a la toma de decisiones mediante la visualización de grandes volúmenes de datos.

Actualmente, la visualización de los datos es vital si se quiere un consumo efectivo de Big Data, por lo que ITI ha centrado sus esfuerzos en desarrollar metodologías de extracción de valor de los datos conforme a la estrategia de la empresa y la aplicación de técnicas de visualización que consideren la enorme cantidad y heterogeneidad de datos disponibles. El resultado de todo ello ha sido el uso y la creación de técnicas y tecnologías que permitan construir soluciones de visualización de datos con las siguientes características:

- **Exploración de datos:** Mediante herramientas diseñadas para descubrir y transformar la manera en la que se toman las decisiones.
- **Análisis rápidos y en tiempo real:** conectando y accediendo de forma rápida a los almacenes de datos mediante tecnologías *in-memory*.
- **Centrado en el usuario final:** visualización y análisis adaptados a la estrategia de la empresa, incluyendo capacidades de análisis colaborativo.
- **Heterogeneidad de los datos:** trabajando con multitud de tipos y fuentes de datos. Desde hojas de cálculo, hasta servicios en la nube, pasando por bases de datos, sistemas Hadoop o contenidos multimedia.
- **Dashboards y Cuadros de Mando inteligentes:** integrando diferentes prácticas de visualización que combinan diversas vistas de datos para tener un conocimiento más profundo.

## BIBLIOGRAFÍA Y FUENTES DE INTERÉS

A continuación se indican las principales fuentes de información utilizadas para la elaboración del presente informe:

- [1] **IBM Institute for Business Value, Escuela de Negocios Saïd en la universidad Universidad de Oxford.** *Analytics: el uso de big data en el mundo real.* 2014. [En línea] Disponible en: [http://www-05.ibm.com/services/es/gbs/consulting/pdf/El\\_uso\\_de\\_Big\\_Data\\_en\\_el\\_mundo\\_real.pdf](http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf)
- [2] **OBS Business School.** *Big Data 2015.* Marzo 2015. Disponible a través de solicitud en: <http://www.obs-edu.com/noticias/estudio-obs/en-2020-mas-de-30-mil-millones-de-dispositivos-estaran-conectados-internet/>
- [3] **Warsaw Institute for Economic Studies.** *Big & Open data in Europe, A growth engine or a missed opportunity.* 2014. [En línea] Disponible en: [http://www.bigopendata.eu/wp-content/uploads/2014/01/bod\\_europe\\_2020\\_full\\_report\\_singlepage.pdf](http://www.bigopendata.eu/wp-content/uploads/2014/01/bod_europe_2020_full_report_singlepage.pdf)
- [4] **H. Gilbert Miller, Peter Mork.** *From Data to Decisions: A Value Chain for Big Data.* ", IT Professional, vol.15, no. 1, pp. 57-59, Jan.-Feb. 2013, doi:10.1109/MITP.2013.11. [En línea] Disponible en: [http://www.fh-schmalkalden.de/Englmeier-p-790/\\_/ValueChainBigData.pdf](http://www.fh-schmalkalden.de/Englmeier-p-790/_/ValueChainBigData.pdf)
- [5] **HP.** *From big data to knowledge: value chain for CSPs,* Business White paper. Noviembre 2012. [En línea] Disponible en: <http://www.vertica.com/wp-content/uploads/2013/02/From-Big-Data-to-Knowledge-Value-Chain-for-CSPs-4AA4-3407ENW1.pdf>
- [6] **BDVA.** *European Big Data Value Strategic Research & Innovation Agenda.* Enero 2015. [En línea] Disponible en: <http://bigdatavalue.eu/index.php/downloads/finish/3-big-data-value/72-big-data-value-strategic-research-and-innovation-agenda/0>
- [7] **McKinsey Global Institute.** *Big data: The next frontier for innovation, competition, and productivity.* Junio 2011. [En línea] Disponible en: [http://bigdatawg.nist.gov/MGI\\_big\\_data\\_full\\_report.pdf](http://bigdatawg.nist.gov/MGI_big_data_full_report.pdf)
- [8] **Rock Health.** *Big Data in Digital Health.* 2012. [En línea] Disponible en: <http://www.slideshare.net/RockHealth/rock-report-big-data>
- [9] **Instituto Valenciano de Tecnologías Turísticas (Invat.tur), Territorio Creativo.** *Big Data: retos y oportunidades para el turismo.* Febrero 2015. [En línea] Disponible en: <http://www.slideshare.net/invattur/estudio-big-data-retos-y-oportunidades-para-el-turismo>
- [10] **Amadeus IT Group.** *At the Big Data Crossroads: turning towards a smarter travel experience.* Junio 2013. [En línea] Disponible en: [http://www.bigdata.amadeus.com/assets/pdf/Amadeus\\_Big\\_Data.pdf](http://www.bigdata.amadeus.com/assets/pdf/Amadeus_Big_Data.pdf)
- [11] **ITI.** *Las Tecnologías de la Información y las Comunicaciones para la Industria del Futuro.* Noviembre 2014. [En línea] Disponible en: <http://www.coiicv.org/publicaciones/download/23-monografias/308-las-tecnologias-de-la-informacion-y-las-comunicaciones-para-la-industria-del-futuro>
- [12] **Economist Intelligence Unit.** *Manufacturing and the Data Conundrum.* Julio 2014. [En línea] Disponible en: [http://www.economistinsights.com/sites/default/files/Manufacturing\\_Data\\_Conundrum\\_Jul14.pdf](http://www.economistinsights.com/sites/default/files/Manufacturing_Data_Conundrum_Jul14.pdf)
- [13] **García-Montalvo, J.** *El impacto del Big Data en los servicios financieros.* 2014. Papeles de Economía Española, nº extraordinario, (Número monográfico sobre "Nuevos negocios bancarios"), pp. 43-58. [En línea] Disponible en: [http://www.econ.upf.edu/~montalvo/wp/big\\_data\\_banking](http://www.econ.upf.edu/~montalvo/wp/big_data_banking)

[v4.pdf](#)

[14] **ITI**. *Smart Environments: Las TIC en las Ciudades Inteligentes*. Octubre 2011. [En línea] Disponible en: [https://observatorio.iti.upv.es/media/managed\\_files/2011/10/06/Informe\\_Breve\\_SMARTCITIES.pdf](https://observatorio.iti.upv.es/media/managed_files/2011/10/06/Informe_Breve_SMARTCITIES.pdf)

[15] **Chicago Council on Global Affairs**. *The Emerging Power of Big Data: The Chicago Experience*. Junio 2014. [En línea] Disponible en: <http://newamerica.net/sites/newamerica.net/files/policydocs/Emerging%20Power%20of%20Big%20Data%20-%202014.pdf>

[16] **EY**. *Sustaining digital leadership*. Marzo 2014. [En línea] Disponible en: [<http://www.ey.com/GL/en/Industries/Technology/EY-Sustaining-digital-leadership>



**ITI**  
INSTITUTO TECNOLÓGICO  
DE INFORMÁTICA

Cofinanciado por:

