



OBS Business
School

Big Data vs. Huge Data: más similitudes que diferencias

Óscar Quero

Profesor en OBS Business School

Junio, 2022

Partners Académicos:



UNIVERSITAT DE
BARCELONA

UIC
barcelona

OBSbusiness.school

Autor



Oscar Quero

Profesor en OBS Business School



Óscar Quero es un profesional con más de 20 años de experiencia, durante los cuales ha podido aportar valor a diferentes compañías líderes de sus sectores, implementando soluciones innovadoras que les han permitido obtener el máximo posible del mayor activo que todas las empresas poseen: la información.

Durante su larga trayectoria profesional, mayoritariamente en cargos de responsabilidad en compañías multinacionales, Quero ha liderado multitud de proyectos centrados en los datos, colaborando con los diferentes departamentos de las compañías.

Su experiencia con negocio, unido a un conocimiento técnico de multitud de herramientas y tecnologías, le han convertido en un profesional de referencia en el sector, habiendo participado como ponente en eventos y actividades durante los últimos años.

Actualmente, desempeña su labor como Business Information Manager de todo internacional (Europa, Asia, Pacífico y Australasia) para la compañía Teladoc Health, líder en telemedicina a nivel mundial, siendo responsable de todo el proceso de innovación tecnológica en torno a los datos, la información y el conocimiento.



Índice

Capítulo 1	Introducción	05
	¿Qué es Big Data?	06
	Ejemplos de aplicación de Big Data	09
Capítulo 2	Cuándo se pasa del Big Data al Huge Data	11
	¿Qué es Huge Data?	12
	Los Data Lakes	14
	Ejemplo real de uso de Huge Data	14
Capítulo 3	El Big/Huge Data en la empresa	18
	¿Cuándo se podría/debería implementar?	19
	¿Se necesita disponer de personal con habilidades concretas?	19
	¿Qué departamentos pueden resultar más beneficiados por el uso de este tipo de tecnologías?	21
	¿El uso de Big Data está muy extendido?	22
	Casos de éxito de Big Data: Netflix	23
Capítulo 4	Big/Huge Data y la Ciberseguridad	24
Capítulo 5	Futuro del Big Data/Huge Data	26
Capítulo 6	Conclusiones	27
	Referencias bibliográficas	29



Capítulo 1

¿Qué es Big Data?

1

¿Qué es Big Data?

Cuando hablamos de Big Data, este concepto se refiere a la gestión de grandes volúmenes de datos que se realiza por parte de las diferentes empresas y organismos públicos. Los datos pueden ser de tres tipos:

- **Estructurados.** Son datos con algún tipo de estructura, como los que se obtienen de bases de datos.
- **No estructurados.** Son aquellos que carecen de estructura alguna, como podrían ser las fotos, conversaciones, etc.
- **Semiestructurados.** Son datos que combinan características de los dos anteriores. Ejemplos de ellos serían los que tienen formato HTML, XML o JSON.

Estos datos, para ser considerados como Big Data, deben poseer al menos una de estas características:

- **Volumen.** Cuando se habla de volumen siempre nos referimos a cantidad de datos generados.
- **Variabilidad.** Una de las grandes complejidades con las que las empresas se encuentran es la diversidad en cuanto a la tipología de los datos. Los datos han pasado de ser siempre estructurados a ser semiestructurados o totalmente desestructurados (una foto, por ejemplo).
- **Velocidad.** Los datos deben ser procesados cada vez más rápido, ya que su ciclo de vida es cada vez más corto (pierden valor según el tiempo pasa). Por tanto, se deben extraer, integrar y analizar con rapidez para obtener el máximo valor.

No obstante, aunque tradicionalmente solo se habla de las 3 uves del Big Data (citadas anteriormente), diferentes expertos amplían la lista añadiendo cuatro más, para completar una lista de 7 uves:

- **Veracidad de los datos.** Esta característica se refiere a la fiabilidad que tienen los datos, es decir, es lo que se suele llamar “incertidumbre de los datos”.
- **Viabilidad.** Se refiere a la “capacidad” que los diferentes entes públicos o privados tienen para utilizar de forma eficaz los datos recabados.
- **Visualización de los datos.** De nada vale obtener los datos e integrarlos si no podemos visualizarlos de forma que se los pueda explotar de una manera adecuada para obtener el máximo de ellos.

- **Valor de los datos.** No todos los datos son igual de importantes ni tienen el mismo valor. Las compañías deben analizar los datos para integrar solo aquellos que ayudan a la toma de decisiones.

El concepto de Big Data, aunque *a priori* pueda parecer moderno, tiene ya muchos años de historia. Sectores como la banca o los seguros han utilizado tradicionalmente grandes volúmenes de datos para sus operaciones. No obstante, el cambio se ha producido principalmente por estos factores:

- Se dispone de nuevas fuentes de datos de alto valor y complejidad: redes sociales, imágenes, etc.
- Auge de IoT (Internet of Things).
- Aparición de nuevas tecnologías para el almacenamiento y gestión de los datos.
- Expansión de su uso a todos los sectores económicos.



Si analizamos retrospectivamente los diferentes hitos que se relacionan con el Big Data, los más relevantes podrían ser:

- **1956:** Aparición de la memoria virtual.
- **1961:** Ley bibliométrica del aumento exponencial.
- **1975:** La comunicación bidireccional.
- **1980:** Ley de Parkinson.
- **1989:** Creación de la World Wide Web.
- **1992:** El primer informe con datos de una base de datos.
- **1995:** Se extiende el uso de la World Wide Web.
- **1997:** Se empieza a plantear el problema de Big Data.
- **2001:** Se definen las 3 uves de Big Data (Gartner).
- **2003:** Publicación por parte de Google del Google File System (GFS) y MapReduce.
- **2006:** Aparece Hadoop.
- **2009:** Aparece Cloudera en el mercado.
- **2011:** Lanzamiento de Hortonworks, que es un competidor de Cloudera.
- **2012:** Se empieza a aplicar tecnología Big Data en campañas políticas como la de Barak Obama (EE.UU.).
- **2013:** Aumenta exponencialmente el interés a nivel mundial del Big Data, creciendo el número de publicaciones especializadas (más de 250 libros en Amazon).
- **2014:** Nace el concepto de “location intelligence”, consistente en la gestión y análisis de los datos de geolocalización.
- **2015:** Nacimiento de las *smart cities*.
- **2020:** IoT y la masificación en la generación de datos multidispositivo.
- **2021:** Uso de Big Data para monitorizar la pandemia del COVID-19.
- **2022:** La ciberseguridad empieza a ser un área de inversión de las empresas en Big Data, principalmente por el incremento de los ataques por la guerra de Ucrania.

Como puede verse de forma clara, Big Data es un concepto con un largo recorrido que ha ido evolucionando con el paso del tiempo, aunque la explosión de IoT y el desarrollo de nuevas tecnologías para la gestión de los datos de forma rápida y eficiente ha provocado que compañías de diferentes tamaños y sectores puedan embarcarse en proyectos de este tipo.

2

Ejemplos de aplicación de Big Data

—> Big Data e inteligencia artificial

Cada año que pasa, la cantidad de datos que se genera se duplica, lo que está llevando a las empresas a un escenario complicado, ya que no pueden gestionar los datos que generan y, por lo tanto, tomar las mejores decisiones basadas en ellos (*data-driven companies*).

La utilización de algoritmos y sistemas Machine Learning, que se ven ampliamente potenciados por el enorme volumen de datos al que pueden acceder, permite definir sistemas “inteligentes” que estén constantemente aprendiendo a procesar los datos para obtener el máximo valor posible de ellos.

Si exploramos la implantación de esta tecnología a nivel empresarial, podemos encontrar diferentes ejemplos o casos de éxito:

- **Google y su IA utilizada en medicina.** La compañía americana absorbió parte del equipo de una compañía llamada DeepMind, que poseía una aplicación que se dedicaba a la monitorización médica en el Servicio Nacional de Salud de Reino Unido (NHS), y a la que quieren convertir, según sus palabras, en “un asistente con inteligencia artificial para doctores y enfermeras”.
- **Google y su IA utilizada para descifrar textos antiguos.** DeepMind, en colaboración con la Universidad de Oxford, ha logrado descifrar textos griegos antiguos, en algunos casos ilegibles, gracias a la aplicación de inteligencia artificial, arrojando unos resultados excepcionales.
- **Metlife y su uso de IA.** Esta empresa líder del sector asegurador ha desarrollado un proyecto llamado “Artificial Intelligence”, en el que han enriquecido los datos que se manejaban con otros no estructurados (informes médicos, por ejemplo), pudiendo analizar cada interacción en detalle con el máximo de información disponible.



—> **Big Data y Deep Learning**

Deep Learning o aprendizaje profundo es un concepto ligado a las redes neuronales artificiales, por lo que si se quiere poder procesar redes con gran cantidad de nodos se requiere este tipo de tecnología. En este enfoque se utilizan diferentes estructuras que se asemejan al sistema nervioso de un mamífero, organizados en neuronas (nodos) que están especializadas en detectar una serie de características. Los modelos computacionales de Deep Learning “imitan” esos sistemas nerviosos y en la actualidad se enfocan más en las áreas del aprendizaje y el lenguaje.

¿Qué podemos esperar del Deep Learning en el futuro? Lo que actualmente se ha desarrollado son los primeros pasos, aunque ya se está trabajando en otras áreas como las siguientes:

- Memoria semántica
- Razonamiento
- Atención
- Motivación
- Emoción

En el futuro se dispondrá de “sistemas cognitivos artificiales” capaces de aprender y captar información para procesarla del mismo modo que lo hace nuestro cerebro.




```
elif operation == "MIRROR Y":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = True  
    mirror_mod.use_z = False  
elif operation == "MIRROR Z":
```

Capítulo 2

Cuándo se pasa del Big Data al Huge Data

1

¿Qué es Huge Data?

Hoy en día todas las empresas tienen puesto el foco en el análisis de los datos, tanto internos como externos, para la toma de decisiones.

Hace años, la gestión de datos era algo manejable por las empresas, pero la evolución en cuanto a cantidad y tipología, unida a la “caducidad” de algunos datos y a la necesidad imperiosa de disponer de información de calidad en tiempo real, hizo que tuvieran que ir un paso más allá para poder trabajar en este nuevo escenario.

Esta situación se ha mantenido durante varios años, pero la llegada del Internet de las cosas (IoT), con volúmenes de generación de datos enormes y multidispositivo (además de ser en muchos casos desestructurados), ha supuesto una nueva “vuelta de tuerca” y se ha tenido que replantear el modo de gestionarlos, ya que según diferentes estudios se estima que el volumen de datos generados anualmente equivale en la actualidad a 2,5 quintillones.

¿Están las tecnologías actuales preparadas para gestionar estos volúmenes de datos y cumplir con las necesidades empresariales en cuanto a velocidad y calidad? La respuesta inicial es no, aunque hay dos tecnologías que sí lo permiten: Inteligencia Artificial (IA) y Deep Learning. El concepto de Huge Data se refiere tanto al enorme volumen de datos, en muchos casos inmanejable, como a la exploración de nuevos enfoques y tecnologías para su gestión eficiente.

¿Cómo se relacionan Big Data, Machine Learning y Deep Learning? En sí mismas son tecnologías complementarias, no excluyentes, por lo que cada una tiene su propio campo de actuación:

- **Big Data:** Se encarga de extraer y procesar los datos de diferentes orígenes para ponerlos a disposición de los diferentes algoritmos de Machine Learning y Deep Learning.
- **Machine Learning:** Trabaja con los datos procesados por la plataforma Big Data para obtener conocimientos para negocio y automatizar ciertas tareas (el sistema aprende).
- **Deep Learning:** Utiliza los datos procesados por Big Data y aprende a altos niveles de profundidad, focalizándose en tareas de alta complejidad.

La llegada del IoT, a lo que se añade el uso de sensores en almacenes, fábricas y otros entornos empresariales (incluyendo sensores de calor y de control de aforo, tan importantes ahora por la pandemia de COVID-19), unido al aumento en la velocidad de generación de los datos, está obligando a las compañías a transformarse y adaptarse para no morir por exceso de datos. Las infraestructuras, sistemas y procesos no suelen estar preparados para absorber este tsunami de datos y seguir proveyendo de datos fiables a las áreas encargadas de la toma de decisiones.



Los sistemas actuales Big Data no están preparados para absorberlo y procesarlo todo con la velocidad que las empresas necesitan, por lo que es necesario implementar sistemas Machine Learning y Deep Learning que permitan procesar la información, obtener el valor de cada dato y proveer de información de calidad a los diferentes usuarios para favorecer la toma de decisiones.

¿Es esto sencillo? La verdad es que son proyectos con un alto nivel de complejidad, y que deben acometerse en fases, siguiendo el mismo principio que se sigue en Business Intelligence (de más superficial a más profundo). Por tanto, todo proyecto de este tipo debe comenzar con la implementación de una infraestructura y arquitectura Big Data que, una vez esté madura, nos permita comenzar a trabajar en el primer algoritmo Machine Learning, que es un tipo de algoritmo guiado (no toma decisiones por sí mismo).

Una vez la compañía se encuentre más madura en cuanto a este tipo de sistemas, se puede proponer avanzar a Deep Learning, diseñando una red neuronal. Esto supone un gran esfuerzo para la compañía, ya que la complejidad no está solo en diseñar este sistema sino en gestionar el cambio cultural interno, puesto que, a diferencia de los algoritmos de Machine Learning, estos sistemas cognitivos aprenden y toman decisiones por sí solos.

El secreto del éxito de estos proyectos es el análisis previo de los datos para definir qué aporta valor en cada momento. Este análisis no debe ser algo puntual, ya que las necesidades van cambiando y debemos adaptar los procesos de forma continua.

2

Los Data Lakes

Si ya con los sistemas Big Data los Data Lakes empezaron a tener más relevancia, con Huge Data pasan a ser una necesidad. El enorme volumen de datos que se extrae de los diferentes sistemas y dispositivos es importante almacenarlo en una ubicación en sus formatos de origen, ya que los indicadores analizados en los diferentes momentos o los datos considerados “de valor” pueden cambiar según evolucionan los escenarios, por lo que es necesario asegurar que se dispone de toda la información.

¿Qué permite un Data Lake? Implementar un sistema de almacenamiento de este tipo tiene muchas ventajas:

- Posibilidad de guardar todos los datos en bruto, es decir, sin procesamiento de ningún tipo.
- Capacidad de disponer de los datos que necesitan los diferentes perfiles de usuarios.
- Facilidad de adaptar el sistema analítico en función de las necesidades.

Por tanto, y dado que el coste de almacenamiento ha decrecido considerablemente, es recomendable implementar este tipo de tecnologías para flexibilizar la plataforma analítica y poder dar servicio a cualquier necesidad, siempre y cuando los datos existan.

3

Ejemplo real de uso de Huge Data

2020 va a ser siempre recordado por la pandemia mundial de COVID-19, que ha provocado una situación muy grave en todos los países y que tendrá un impacto enorme a nivel social y económico, ya que no solo va a impactar en aspectos como el consumo, sino que está generando un cambio en las conductas individuales. Todos los habitantes, como consecuencia del miedo y la incertidumbre, hemos frenado el consumo y la socialización, dejando de realizar actividades de ocio como ir al cine o a un restaurante, sustituyéndolas por una mayor vida familiar y los pedidos a domicilio, además de incrementar el consumo de servicios como Netflix o HBO (en sustitución del cine).

¿Qué nos ha demostrado esta pandemia? Que en un mundo tan globalizado como el actual, donde la movilidad es cada vez más sencilla, es muy difícil contener la dispersión de un virus, más aún si tiene características que lo pueden hacer indetectable como el actual coronavirus (85 % de asintomáticos). Los métodos de contención habituales han demostrado ser muy ineficaces, principalmente por la lentitud en la respuesta y la falta de contundencia por parte de los organismos sanitarios, por lo que se ha

constatado que ante situaciones como esta es muy necesario disponer de información de calidad de forma muy rápida, para de este modo poder conocer la situación real y tomar decisiones.

Por todo lo explicado anteriormente, unido al hecho que la diseminación del virus haya sido tan rápida (alto nivel de contagio), todos los organismos públicos estatales y locales se han visto obligados a implementar de forma urgente sistemas que permitan conocer la situación actual en diferentes indicadores, de cara a implementar medidas para contener el virus. Uno de estos indicadores es la movilidad de los habitantes.

Por tanto, esta pandemia ha traído muchas desgracias, pero también ha dejado cosas buenas, como la demostración práctica de la aplicabilidad y beneficios de la tecnología Big Data/Huge Data para analizar indicadores clave para la toma de decisiones, donde el tiempo de reacción y la calidad de los datos es clave.

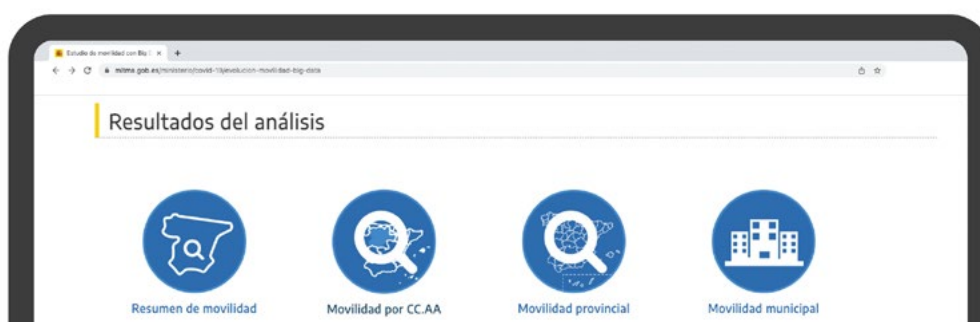
El siguiente ejemplo es de una plataforma del Ministerio de Transportes, Movilidad y Agenda Urbana, que realiza un *Análisis de la movilidad en España durante el estado de alarma*. Los datos que utiliza este sistema provienen de las antenas para móviles, dispersas por todo el territorio nacional, de donde se extraen los datos de posicionamiento de los diferentes dispositivos para su análisis por cualquier usuario (son datos accesibles). Veamos más en detalle la solución.

Lo primero que nos encontramos, aparte de la explicación metodológica, es el acceso a diferentes niveles de análisis:

- Resumen de movilidad
- Movilidad por CC. AA.
- Movilidad provincial
- Movilidad municipal

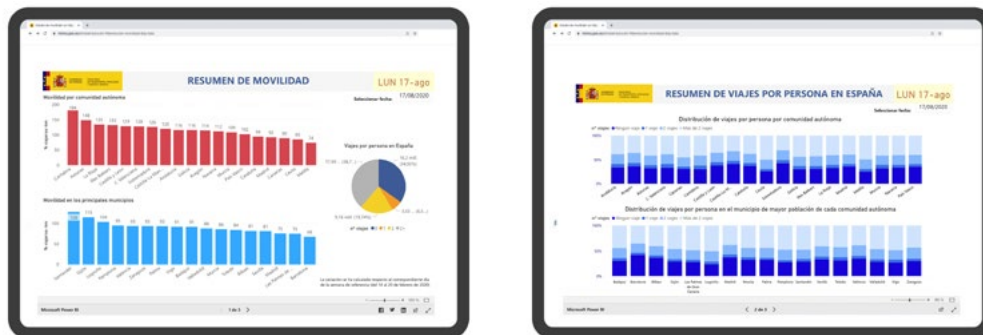
Figura 01 → RESUMEN DE MOVILIDAD

Fuente: Ministerio de Transportes,
Movilidad y Agenda Urbana

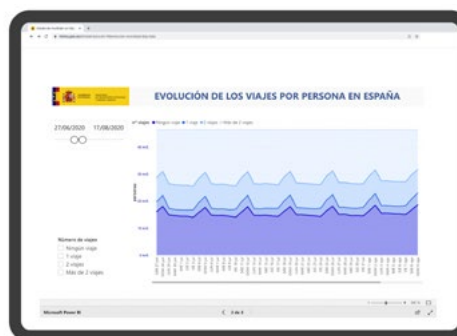


Si accedemos a la opción de “Resumen de movilidad”, entramos en la aplicación analítica (desarrollada en Power BI), que nos muestra tres hojas de análisis:

- Resumen de movilidad

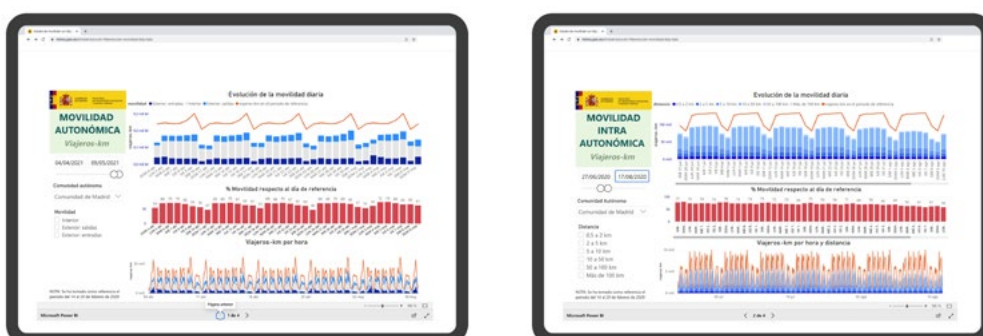


- Evolución de los viajes por persona en España.



Si pasamos al análisis de la información por comunidad autónoma, el enfoque cambia, volviéndose más granular:

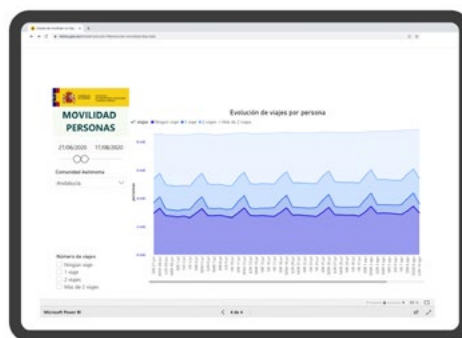
- Movilidad autonómica



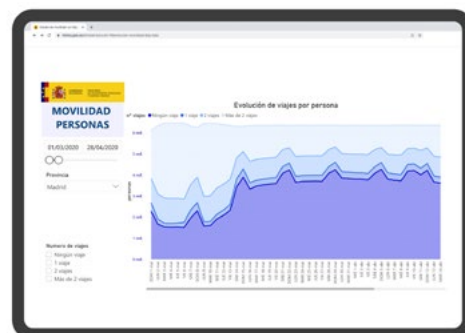
- Movilidad inter-autonómica.



- Movilidad personas



Estos mismos análisis que se pueden ver a nivel comunidad autónoma se han implementado también a nivel provincial y municipal. Estos son algunos ejemplos:



Este es un ejemplo claro de cómo el uso de la tecnología para gestionar Huge Data permite obtener información de calidad que ayuda a la toma de decisiones, detectando situaciones tanto positivas como negativas.



Capítulo 3

El Big/Huge Data en la empresa

1

¿Cuándo se podría/debería implementar?

Implementar o desarrollar un proyecto Big Data nunca debería ser una cuestión de dinero ni la decisión debería basarse únicamente en la necesidad de “estar a la última”, ya que no todas las empresas están preparadas para realizar proyectos de este tipo.

Un aspecto importante que no debemos obviar y que es una de las bases de este tipo de proyectos es el volumen de datos. No todas las empresas tienen el suficiente volumen de datos como para aprovechar la potencia de una herramienta de Big Data. Es cierto que muchas empresas lo implementan de todos modos, pero al final esto se convierte en un arma de doble filo, ya que probablemente no puedan recuperar nunca la inversión realizada, que en muchos casos es muy elevada. En una frase más coloquial, “están matando moscas a cañonazos”.

Otro de los aspectos no menores a tener en cuenta, y al que muchas compañías no dan importancia, es la madurez empresarial. Según CMMI, que es un modelo que mide la madurez de las organizaciones, una compañía estaría preparada para acometer un proyecto de este tipo cuando se encuentre en un nivel tres de madurez (idealmente el cuatro), ya que debe tener procesos sistemáticos y proactivos. Compañías con niveles de madurez inferiores también podrían embarcarse en este tipo de iniciativas, pero la probabilidad de fracaso se incrementaría exponencialmente.

2

¿Se necesita disponer de personal con habilidades concretas?

La respuesta sonará a ambigua, pero es: “depende”. Depende de lo que se quiera hacer con Big Data en la organización, ya que esto marcará el tipo de habilidades que el equipo de BI, ya sea interno o externo, deberá poseer en su conjunto para poder gestionar el servicio.

Ante todo, es importante conocer las habilidades que se necesitan para gestionar proyectos de este tipo, que son las siguientes:

- **Habilidades y conocimientos de programación.** Debemos tener recursos capacitados en lenguajes como Python (muy de moda actualmente) y con conocimiento de los marcos Apache Spark y MapReduce.
- **Conocer el “aprendizaje automático”.** La estadística es clave en el máximo aprovechamiento de los datos y conocerla es fundamental para poder desarrollar los diferentes modelos predictivos.
- **Conocimientos de minería de datos.** Es importante tener experiencia en herramientas de minería de datos, ya que esta es una de las técnicas que ayuda a incrementar el valor que los datos nos aportan.

- **Perfil analítico.** Si el equipo, en su conjunto, no tiene un perfil realmente analítico y conocimientos/experiencia en análisis cuantitativo/cualitativo y descriptivo/predictivo/prescriptivo, será muy complicado que las iniciativas de Big Data lleguen a buen puerto.
- **Conocimiento y experiencia con alguna herramienta de visualización de datos.** De nada vale trabajar bien los datos, si no los mostramos de una forma sencilla e intuitiva. Necesitamos disponer siempre de personas con habilidades de diseño, que sepan cuál es la mejor manera de mostrar la información

En base a estas habilidades, estos serían algunos ejemplos de roles organizativos:

- **Científicos de datos.** Es de los perfiles más complejos y amplios, ya que se encargan de recopilar, analizar, gestionar, estructurar e interpretar los datos, siendo expertos en identificar “los patrones de los datos”, además de las relaciones inherentes y las tendencias existentes.
- **Arquitectos de datos.** Son los que “piensan” en cómo debería ser toda la arquitectura de sistemas y la estructura de datos, y apoyan a los científicos de datos.
- **Analistas.** Son los “investigadores de los datos”. Se encargan de entender la información y explorarla, para detectar patrones o tendencias que comunicar a negocio.
- **Desarrolladores.** Expertos en uno o más lenguajes de programación aplicables a Big Data (Python, Scala, etc.).
- **Especialistas Big Data.** Son los que más saben acerca de cómo gestionar y tratar los grandes volúmenes de datos, en muchos casos muy complejos.

¿Realmente es necesario disponer de estos perfiles en la compañía? En un mundo ideal, sí, pero la realidad empresarial nos lleva a estructuras donde no existen perfiles tan puros, sino mezcla de perfiles (o más bien de conocimientos y habilidades).



3

¿Qué departamentos pueden resultar más beneficiados por el uso de este tipo de tecnologías?

Aunque realmente toda la organización se beneficiará de cualquier implementación exitosa de Big Data, hay ciertas áreas donde el impacto es más alto. Estas son las siguientes:

- Dirección estratégica
- Marketing y Ventas
- Innovación
- Recursos Humanos

Para cada una de estas áreas, se pueden identificar diferentes beneficios:

- **Dirección estratégica.** Gracias a Big Data, este departamento podrá entender las diferentes dinámicas del mercado y cómo funciona la competencia, pudiendo de esta manera reaccionar de forma rápida ante los riesgos y poder aprovechar las potenciales oportunidades.
- **Marketing y Ventas.** Esta es el área por antonomasia que se beneficia de verdad de implementar proyectos de este tipo. Gracias a monitorizar de forma constante las redes sociales, entendiendo cómo piensan y actúan los potenciales clientes, además de seguir las acciones de la competencia, pueden actuar de forma rápida y adaptar la “táctica” para lograr o superar los objetivos planteados.
- **Innovación.** Las áreas de I+D+I son departamentos que se pueden aprovechar mucho del análisis de grandes volúmenes de datos. Pongamos como ejemplo el análisis de los resultados obtenidos de un estudio clínico. Este tipo de estudios contiene multitud de información, ya que hay información sobre los pacientes y los diferentes análisis realizados (y cuestionarios realizados), por lo que un análisis de todo es complejo. Gracias a Big Data se pueden establecer relaciones entre los datos, normalmente de causalidad, lo que permite a los investigadores entender qué está pasando y poder sacar conclusiones apoyadas en datos concretos. Además, el Big Data puede utilizarse para otros fines, como el seguimiento de todo lo que se publica en las diferentes revistas científicas, lo que permitirá estar al tanto de todo lo que puede afectar a su día a día.
- **Recursos Humanos.** Es el área que más podría beneficiarse de implementar proyectos Big Data, dado el tipo de información que gestiona y el volumen de datos e indicadores. Una de las utilidades de este tipo de proyecto puede ser conocer mejor a los empleados, entendiendo su modo de trabajo y cómo utilizan los diferentes dispositivos, pudiendo medir aspectos críticos para este departamento como la productividad.

4

¿El uso de Big Data está muy extendido?

Para poder entender si realmente las empresas “confían” en los proyectos de este tipo, se ha realizado un análisis de los datos estadísticos recogidos en el INE (Instituto Nacional de Estadística) sobre Big Data:

	Total Empresas	Industria	Construcción	Servicios
Total nacional				
% de empresas que analizaron Big Data	11,05	8,08	5,03	13,81
% de empresas que analizaron Big Data por tipo de fuente: datos de la propia empresa con sensores o dispositivos inteligentes (1)	25,09	41,07	16,68	21,87
% de empresas que analizaron Big Data por tipo de fuente: datos por geolocalización a partir de dispositivos portátiles (1)	54,06	39,88	76,64	55,55
% de empresas que analizaron Big Data por tipo de fuente: datos generados por medios sociales (1)	45,30	37,04	25,83	49,12
% de empresas que analizaron Big Data por tipo de fuente: otras fuentes de Big Data (1)	27,17	27,88	15,52	28,05

Como se puede ver, en España no hay un porcentaje grande de empresas que usen Big Data, siendo mayoritariamente las empresas del sector servicios las que lo hacen. Los datos que mayoritariamente son analizados con este tipo de tecnologías son los de geolocalización y de las redes sociales, lo que es en principio lógico por la necesidad de las compañías de gestionar el alto volumen de datos generado por los dispositivos.

Si analizamos los datos por tamaño de empresa:

	Total	De 10 a 49	De 50 a 249	De 250 y más
Total nacional				
% de empresas que analizaron Big Data	11,05	9,44	17,49	29,40
% de empresas que analizaron Big Data por tipo de fuente: datos de la propia empresa con sensores o dispositivos inteligentes (1)	25,09	18,30	38,79	52,37
% de empresas que analizaron Big Data por tipo de fuente: datos por geolocalización a partir de dispositivos portátiles (1)	54,06	55,53	50,10	51,07
% de empresas que analizaron Big Data por tipo de fuente: datos generados por medios sociales (1)	45,30	45,62	42,78	49,57
% de empresas que analizaron Big Data por tipo de fuente: otras fuentes de Big Data (1)	27,17	23,92	31,92	45,60

Está claro que el uso está más focalizado en grandes empresas, posiblemente por la inversión necesaria para proyectos de este tipo, pero llama la atención el alto nivel de implementación en PYMEs.

En cuanto al uso que suelen darle las empresas, todas apuestan por proyectos de análisis de datos de geolocalización y de datos de redes sociales, lo que demuestra el valor que estos datos tienen para todas las empresas en la actualidad y la dificultad de su gestión con tecnologías no Big Data.

5

Casos de éxito de Big Data: Netflix

Netflix es una de las empresas más innovadoras en este tipo de tecnología, ya que aplica un sistema que recomienda contenidos en base a tecnología de Machine Learning. Esta plataforma recaba una gran cantidad de los usuarios y analiza los hábitos de consumo gracias a un algoritmo, proponiendo contenidos similares a otros usuarios con perfil similar. Esto permite mejorar la experiencia de los usuarios y ofrecer contenidos “frescos” de forma constante, lo que logra mantener el interés en lo que la plataforma puede ofrecer.

¿Cómo funciona este algoritmo? El sistema se encarga de captar toda la información referente a:

- Uso del servicio
- Usuarios con hábitos y perfil similares
- Información sobre los contenidos actuales de la plataforma
- En qué horarios usan los clientes el servicio
- Qué dispositivos se utilizan
- Tiempo de uso

En base a esta información, el sistema prepara una recomendación de contenidos que ordena dentro de cada una de las filas, y se ordenan las filas para que la experiencia del usuario sea la mejor posible, estando lo más recomendado arriba a la izquierda y lo menos abajo a la derecha.



Capítulo 4

Big/Huge Data y la Ciberseguridad





El Big Data es un gran aliado de las compañías, ya que puede ayudar a protegerlas ante los posibles ataques que pueden sufrir, detectando las vulnerabilidades de forma rápida y eficiente. Pero, ¿cómo puede ayudar de forma más concreta? Hay diferentes modos:

- **Creando modelos.** El uso de modelos predictivos puede ayudar a generar alertas de posibles amenazas antes de que realmente se produzcan, pudiendo protegerse la empresa de forma anticipada. Normalmente, con estos modelos se utilizan técnicas de inteligencia artificial y aprendizaje automático.
- **Monitorizando los sistemas.** Se pueden crear diferentes reglas que, en base al análisis de toda la información que generan los sistemas, nos avise de posibles eventos de seguridad desde el primer momento.
- **Detectar los posibles ataques.** Con Big Data se pueden implementar sistemas automatizados que ayuden a la detección precoz de intrusos, detectando los primeros movimientos de los atacantes. Esto permitirá evitar que los problemas generados sean más graves.
- **Analizar ataques anteriores.** Por último, estos sistemas ayudan también a analizar todos los datos que se obtuvieron antes/durante/después de un ataque, pudiendo entender qué es lo que sucedió y los motivos, además de la actuación realizada, ayudando al proceso de mejora continua.

Una de las tecnologías que, junto al Big Data, ayudará a mejorar las soluciones de Ciberseguridad en las compañías es la Inteligencia Artificial (IA). Gracias a esta tecnología, los propios sistemas de las compañías pueden predecir futuros ataques, al estar preparados los algoritmos para trabajar con información de tipos diversos y al disponer de GPUs más potentes en los servidores.

Hay una serie de aspectos que deben tenerse en cuenta para que una implementación de este tipo de proyectos de Ciberseguridad sea un éxito:

- Los datos deben tener la calidad suficiente y sean consistentes.
- Los algoritmos deben entrenarse para que entiendan el contexto de la información.
- Todas las variables que se utilicen en los modelos deben estar definidas de forma previa.

Por último, hay que subrayar que, aunque implementemos los algoritmos más avanzados, nada puede actualmente sustituir a la inteligencia humana. Si se quiere que una iniciativa empresarial de este tipo sea un éxito, es crítico combinar tecnología e inteligencia humana, por lo que las compañías deben tener profesionales preparados que se encarguen de optimizar los modelos y de saber cuáles son las preguntas adecuadas a realizar al sistema.



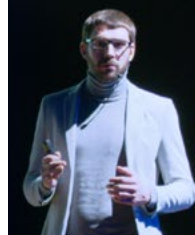
Capítulo 5

Futuro del Big Data/Huge Data



La evolución tecnológica en torno al Big Data lleva una velocidad enorme y no paran de aparecer nuevas soluciones o enfoques innovadores, dado que esta es actualmente una prioridad para las empresas. No obstante, se pueden identificar una serie de áreas de desarrollo:

- **“Datos oscuros” o Dark Data.** Esta tipología de datos, también llamados “Dusty Data”, son aquellos datos de origen no digital que han sido descartados inicialmente por su falta de valor. Hay diferentes tipos de “Dark Data”:
 - » Los datos que no se están recopilando en la actualidad.
 - » Los datos que, aunque se recopilan, es difícil acceder a ellos en tiempo y forma.
 - » Los datos que son recopilados y están disponibles, pero no se utilizan en todo su potencial.
- **Análisis cuántico.** Este tipo de análisis, utilizando ordenadores cuánticos, permitirá procesar grandes volúmenes de datos a velocidades enormes e implementar proyectos tan ambiciosos como el Human Brain Project (HBP), patrocinado por la Unión Europea, que busca replicar el funcionamiento del cerebro humano con tecnología.
- **Big Data en el sector salud.** El sector salud es uno de los focos de los proyectos Big Data. Siguiendo con el proyecto mencionado en el punto anterior, otro objetivo que se han marcado es la recopilación de toda la información de pacientes proveniente de hospitales y clínicas para mejorar los tratamientos prescritos a los pacientes. Es un enfoque muy innovador que ayudará tanto a pacientes como a los entes sanitarios, afinando los tratamientos y las prescripciones asociadas.





Capítulo 5

Conclusiones



- Big Data se encuentra en un nuevo momento de evolución. El momento de la gestión de los grandes volúmenes de datos es parte del pasado, y las compañías deben enfocarse en implementar nuevos sistemas que permitan gestionar volúmenes de datos con incrementos exponenciales en tiempo y forma.
- No todos los datos deben integrarse, ya que no todo tiene el mismo valor en cada momento. Las compañías deben estar continuamente analizando el “valor de los datos” para integrar solo aquellos que ayudan a la toma de decisiones empresariales. Pero este análisis nunca debe ser algo puntual, ya que los escenarios empresariales son dinámicos y las necesidades cambian.
- Aunque no todos los datos aporten valor, es clave no perderlos, ya que el indicador del valor que aportan puede variar en el futuro. Por tanto, es clave hoy en día disponer de al menos un Data Lake para almacenarlo todo, con vistas a su posible uso en el medio y/o largo plazo.
- Los sistemas actuales de Big Data no están preparados para manejar Huge Data, y se deben apoyar en algoritmos que ayuden al manejo de la información, simplificando la detección de *insights*. Tanto Machine Learning como Deep Learning son tecnologías clave en el manejo de estos enormes volúmenes de datos.
- Los causantes de este incremento masivo en el volumen de datos y, por tanto, de la aparición de Huge Data como concepto son principalmente los dispositivos IoT, es decir, aquellos que están constantemente emitiendo información, y elementos como los sensores, que cada vez más se introducen en fábricas, almacenes y tiendas. Además, la explosión en los próximos años de los coches inteligentes será otro elemento que incrementará de forma notable el volumen de información que deben gestionar las empresas.

- Se está produciendo un uso cada vez mayor de esta tecnología, principalmente para analizar información desestructurada proveniente de redes sociales o geolocalización de dispositivos.
- Es necesario que las compañías cuenten con una serie de habilidades concretas para gestionar proyectos de este tipo. De lo contrario, todo proyecto de este tipo está destinado al fracaso.
- La ciberseguridad es un área de trabajo donde el Big Data puede demostrar todo su potencial, y las compañías están invirtiendo cada vez más en esta área.
- La aparición de los ordenadores cuánticos es una oportunidad para todos los desarrollos en esta área, ya que mejorarán de forma significativa la computación y procesamiento de los datos, siendo cada vez más rápidos y eficientes. Esto será clave en determinados sectores, como el médico, ya que ayudará a la implementación de sistemas realmente “inteligentes” para mejorar los diagnósticos y las prescripciones de tratamiento.
- Por último, no se puede obviar uno de los desarrollos donde Huge Data es clave: las redes neuronales artificiales. Proyectos como el Human Brain Project (HBP), patrocinado e impulsado por la Unión Europea, van a significar un antes y después del Big Data, y llevará la tecnología de dato a un nuevo escenario que nos abrirá nuevas y novedosas posibilidades. ¿Estáis preparados?



Referencias bibliográficas

1. Lostalé, E. (2019, 7 febrero). Qué es huge data y por qué implica IoT, IA y blockchain. Kanlli. <https://www.kanlli.com/estrategia-marketing-digital/que-es-huge-data/>
2. Serrano, A. (2018, 30 agosto). El gran auge del Big Data: repaso desde los inicios hasta hoy (Infografía). Ideas para tu Empresa. <https://ideasparatuempresa.vodafone.es/big-data-desde-los-inicios-hoy/>
3. PowerData, G. (s. f.). Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad. <https://www.powerdata.es/big-data>. Recuperado 16 de mayo de 2022, de <https://www.powerdata.es/big-data>
4. Las 7 V del Big data: Características más importantes. (s. f.). <https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/#:~:text=Las%20caracter%C3%ADsticas%20m%C3%A1s%20importantes%20del,la%20de%20Viabilidad%20y%20Visualizaci%C3%B3n.&text=Valor%20de%20los%20datos,-Tambi%C3%A9n%20de%20manera>. Recuperado 16 de mayo de 2022, de <https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/#:~:text=Las%20caracter%C3%ADsticas%20m%C3%A1s%20importantes%20del,la%20de%20Viabilidad%20y%20Visualizaci%C3%B3n.&text=Valor%20de%20los%20datos,-Tambi%C3%A9n%20de%20manera>
5. Diego Calvo. (2019, 2 abril). Tipos de datos: estructurados, semiestructurados y no estructurados. <https://www.diegocalvo.es/tipos-de-datos-estructurados-semiestructurados-y-no-estructurados/>
6. PowerData, R. (s. f.). Inteligencia Artificial, Machine Learning y Big Data. <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/ia-inteligencia-artificial-y-machine-learning-se-vinculan-al-big-data>. Recuperado 16 de mayo de 2022, de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/ia-inteligencia-artificial-y-machine-learning-se-vinculan-al-big-data>
7. SaluDigital, R. (2018, 16 noviembre). El salto inminente de Google por la IA en medicina real. Saludigital. https://www.consalud.es/saludigital/131/google-apuesta-por-la-ia-en-medicina-real_57044_102.html
8. R. (2019, 23 octubre). La IA de DeepMind descifra textos de la antigua Grecia más rápido y con más precisión que los historiadores. La Vanguardia. <https://www.lavanguardia.com/vida/20191023/471159966282/la-ia-de-deepmind-descifra-textos-de-la-antigua-grecia-mas-rapido-y-con-mas-precision-que-los-historiadores.html>

- 9.** Route to Excellence. (s. f.). <https://www.metlife.es/content/dam/metlifecom/es/PDFs/Revistas/Route-to-Excellence-16.pdf>. Recuperado 16 de mayo de 2022, de <https://www.metlife.es/content/dam/metlifecom/es/PDFs/Revistas/Route-to-Excellence-16.pdf>
- 10.** Lazcano, R. (s. f.). Big data, machine learning y deep learning: conceptos y diferencias. <https://blog.enzymeadvisinggroup.com/big-data-machine-learning>. Recuperado 16 de mayo de 2022, de <https://blog.enzymeadvisinggroup.com/big-data-machine-learning>
- 11.** Arrabales, R. (2016, 28 octubre). Deep Learning: qué es y por qué va a ser una tecnología clave en el futuro de la inteligencia artificial. Xataka. <https://www.xataka.com/robotica-e-ia/deep-learning-que-es-y-por-que-va-a-ser-una-tecnologia-clave-en-el-futuro-de-la-inteligencia-artificial>
- 12.** Saiz, Á. (s. f.). Business Insights: ¿qué significan? ¿qué aportan? ¿cómo obtenerlos? <https://blog.enzymeadvisinggroup.com/business-insights>. Recuperado 16 de mayo de 2022, de <https://blog.enzymeadvisinggroup.com/business-insights>
- 13.** Análisis de la movilidad en España durante el Estado de Alarma | Ministerio de Transportes, Movilidad y Agenda Urbana. (s. f.). <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>. Recuperado 16 de mayo de 2022, de <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>
- 14.** Maroto, C. (2020, 22 agosto). El salto del Big Data al Huge Data | Harvard Deusto las revistas. El salto del Big Data al Huge Data. <https://www.harvard-deusto.com/el-salto-del-big-data-al-huge-data>
- 15.** ¿Qué aporta el Data Lake al Big Data? (2017, 25 julio). CICE. <https://www.cice.es/noticia/data-lake-big-data/>
- 16.** Martín, I. (2018, 5 noviembre). Futuro del Big Data – ¿Qué nos espera? <https://blogs.deusto.es/master-informatica/futuro-del-big-data-que-nos-espera/>. <https://blogs.deusto.es/master-informatica/futuro-del-big-data-que-nos-espera/>
- 17.** Banafa, A. (2018, 11 julio). Understanding Dark Data. OpenMind. <https://www.bbvaopenmind.com/en/technology/digital-world/understanding-dark-data/>
- 18.** Hernández, M. (s. f.). Uso del Big Data en el modelo digital más visionario del cerebro humano: el Human Brain Project (HBP). Decideo: Noticias sobre Inteligencia de Negocios, Big Data, Data Science, Data Mining. Recuperado 16 de mayo de 2022, de https://www.decideo.com/Uso-del-Big-Data-en-el-modelo-digital-mas-visionario-del-cerebro-humano-el-Human-Brain-Project-HBP_a1977.html

- 19.** Pozas, J. L. B. (2021, 4 junio). ¿Qué es CMMI? Un modelo para optimizar los procesos de desarrollo. CIO MX. Recuperado 16 de mayo de 2022, de <https://cio.com.mx/que-es-cmmi-un-modelo-para-optimizar-los-procesos-de-desarrollo/>
- 20.** (2021, 17 marzo). » 6 habilidades esenciales de Big Data para conseguir un trabajo de Big Data ». Molinare. Recuperado 22 de mayo de 2022, de <https://molinare.es/2021/03/analisis/6-habilidades-esenciales-de-big-data-para-conseguir-un-trabajo-de-big-data/>
- 21.** Archanco, R. (2015, 16 marzo). El Big Data Intelligence es una actividad de la que pueden beneficiarse casi todas las áreas de una empresa u organización. Pero como en casi todo, siempre habrá alguien que Leer Mas. Papeles de Inteligencia Competitiva. Recuperado 22 de mayo de 2022, de <https://papelesdeinteligencia.com/departamentos-que-se-benefician-del-big-data-intelligence/>
- 22.** INE - Instituto Nacional de Estadística. (s. f.). Análisis de Big Data. INE - Las TIC en las empresas (primer trimestre de 2021) por agrupación de actividad económica (excepto CNAE 56, 64-66 y 95.1) y tamaño de la empresa. Recuperado 4 de junio de 2022, de <https://www.ine.es/jaxi/Tabla.htm?tpx=49864&L=0>
- 23.** INE - Instituto Nacional de Estadística. (s. f.-b). Error. INE - Las TIC en las empresas con 10 o más empleados (primer trimestre de 2021) por agrupación de actividad económica. Recuperado 4 de junio de 2022, de <https://www.ine.es/jaxi/Datos.htm?tpx=49901>
- 24.** Vicente, Z. (2021, 19 marzo). El Big Data y Netflix: te vigilan ¿Lo sabías? Canal Informática y TICS. Recuperado 4 de junio de 2022, de <https://www.inesem.es/revistadigital/informatica-y-tics/big-data-y-netflix/>
- 25.** Cómo funciona el sistema de recomendaciones de Netflix. (s. f.). Centro de ayuda. Recuperado 4 de junio de 2022, de <https://help.netflix.com/es-es/node/100639>
- 26.** Jiménez, J. (2022, 9 enero). Analítica de Big Data: por qué es tan importante en ciberseguridad. RedesZone. Recuperado 4 de junio de 2022, de <https://www.redeszone.net/tutoriales/seguridad/seguridad-informatica-analitica-big-data/>
- 27.** A. (2019, 19 enero). Big Data y Protección de Datos. Big Data y Protección de Datos. Recuperado 4 de junio de 2022, de <https://www.antonioserranoacitores.com/big-data-proteccion-datos/>
- 28.** Ciberseguridad. IA y Big Data: Una inyección de gran potencial a la ciberinteligencia. (2021, 21 diciembre). BABEL Sistemas de Información. Recuperado 11 de junio de 2022, de <https://www.babelgroup.com/es/media/blog/mayo-2021/ciberseguridad-ia-y-big-data-ciberinteligencia>



OBS Business
School

School of **Business
Administration
& Leadership**

School of **Innovation,
& Technology
Management**



De:



Planeta Formación y Universidades