

**El Business Case de Big Data ©
EDICIONES ROBLE, S.L.**

Indice

El Business Case de Big Data	3
I. Introducción	3
II. Objetivos	4
III. Contexto económico	5
3.1 Ley de Moore	5
3.2 Coste de almacenamiento y memoria	7
3.3 Digitalización	8
3.4 Efecto red	9
3.5 Aportación del aprendizaje automático	10
IV. El valor del dato	11
4.1 Impacto del dato (Big Data y Analytics) en las organizaciones	12
4.1.1. Mejora de ingresos / revenue	13
4.1.2. Mejora de márgenes	13
4.1.3. Mejora de eficiencia	13
4.1.4. Competencia	13
4.2 Áreas de gestión del dato	13
V. Tipos de datos desde la perspectiva de negocio	15
VI. Tipos de analítica	17
VII. Puntos clave en la transformación	19
VIII. La organización analítica	21
IX. El Big Data y la empresa	24
9.1 Costes	24
9.2 Agilidad, flexibilidad	25
9.3 Escalabilidad	26
9.4 Competitividad	26
X. Resumen	26
Ejercicios	29
Caso práctico	29
Preguntas	31
Solución	31
Recursos	34
Bibliografía	34
Glosario	34

El Business Case de Big Data

I. Introducción

¹*Disclaimer:* todo el material generado en esta unidad está basado en material previo del autor, por lo que los derechos de propiedad del material base son y se mantienen como propiedad del autor Juan Ramón González.

Estamos afrontando una era de cambio que ninguna generación anterior ha vivido en términos de efectos económicos y sociales. Esta transformación viene dada por el uso masivo de las nuevas tecnologías, que están obligando a que las empresas se transformen digitalmente.

Hace unos años, el problema al que se enfrentaban las empresas era cómo recolectar toda la información de la que disponían sobre su negocio. Ese problema ha sido resuelto y ha dado paso a uno nuevo: ¿cómo explotar y maximizar el uso de esos datos para mejorar el negocio? El nuevo reto, por tanto, consiste en explotar esas ingentes cantidades de información de forma eficiente.

Este creciente volumen de información se debe a varios factores. Por ejemplo:

Web

Toda la información de comportamiento del usuario en medios digitales —cada clic, cada acción— se mide y se recoge y, con ella, es posible conocer el comportamiento digital de los usuarios y, por tanto, hacerlo más eficiente para maximizar el retorno obtenido de ellos.

Redes sociales

Uno de los medios que hoy proporcionan mayor volumen de información son las redes sociales. Todos los días se generan ingentes cantidades de información que, con una correcta explotación, muestran de forma clara las tendencias en opinión y comportamiento colectivo de sus usuarios.

Internet of Things

Hoy en día, la gran mayoría de los dispositivos que tenemos en nuestro hogar tienen una IP. Las ciudades inteligentes, los coches inteligentes... Todos ellos son fuentes de información que vuelcan datos en tiempo real a grandes centros de información, donde, con un procesamiento eficiente, pueden mostrar pautas de comportamiento para optimizar su eficiencia, reducir la tasa de fallos y, en definitiva, optimizar su operativa.

M2M

En este caso, nos referimos a toda la información generada por los *logs* de sistemas e infraestructuras. Un procesamiento inteligente de esta información permite hacer más eficiente la operativa diaria de estos sistemas en tiempo real, calcular indicadores tempranos de fallos y maximizar la utilización de estos sistemas.

Información no estructurada

Hoy en día se estima que más del 80 % de la información disponible es no estructurada, ya sea texto, grabaciones de audio, imágenes, vídeos, etc. Ser capaces de procesar esta información y extraer indicadores de valor abre un abanico de nuevas posibilidades y capacidades para las empresas.

Una forma de entender el Big Data, y más en concreto el aprendizaje automático (inteligencia artificial aplicada al negocio), es como la transformación de la transformación digital.

Actualmente, los directivos de todas las empresas están desarrollando su estrategia digital; en cinco años, los directivos de las grandes corporaciones estarán definiendo una estrategia de inteligencia artificial como máxima prioridad.

II. Objetivos



Los objetivos de esta unidad pueden resumirse en:

- Entender el valor que los datos aportan a las organizaciones.
- Entender el porqué del Big Data y de Analytics hoy en día.

Se analizará, por tanto, la evolución de los sistemas de procesamiento y almacenamiento de datos hasta su estado actual para entender el **porqué de la gran revolución** que han supuesto.

En este aspecto, se analizará:

- El aumento de potencia en los sistemas y cómo ha evolucionado en los últimos años hasta el paradigma Big Data.
- El coste de almacenamiento y memoria. Se estudiará cómo el Big Data abarata los costes de almacenamiento de datos y cómo esto es un requisito básico para la explosión de información de los últimos años.
- Cómo el Big Data y Analytics es una palanca de transformación clave en los procesos de transformación digital de una compañía.
- El valor que aportan los procesos de Business Analytics y aprendizaje automático a una empresa.

Continuaremos luego analizando el valor del dato en las organizaciones y cómo ese dato puede aportar valor en diferentes áreas de negocio.

Se analizarán los tipos de datos, sus características especiales, y se verán los tipos de analítica, tanto desde el punto de vista de qué preguntas de negocio responden como desde la perspectiva de las áreas de negocio a las que aportan valor.

Finalmente, se analizarán diferentes modelos de organización del departamento de analítica en las empresas, sus pros y contras, junto con el grado de madurez necesario para llevarlos a la práctica.

Con todo ello, el alumno tendrá una visión mucho más clara **de por qué, para qué y cómo** se utilizan el Big Data y Analytics como las tres cuestiones clave antes de la adopción de estas técnicas dentro de las organizaciones.

III. Contexto económico

Como hemos comentado en la introducción, estamos en una época de cambio que ninguna otra generación ha vivido anteriormente.

En este apartado, veremos cuáles han sido los cinco factores clave en esta transformación.

3.1 Ley de Moore

La ley de Moore dice que aproximadamente cada dos años se duplica el número de transistores en un microprocesador.

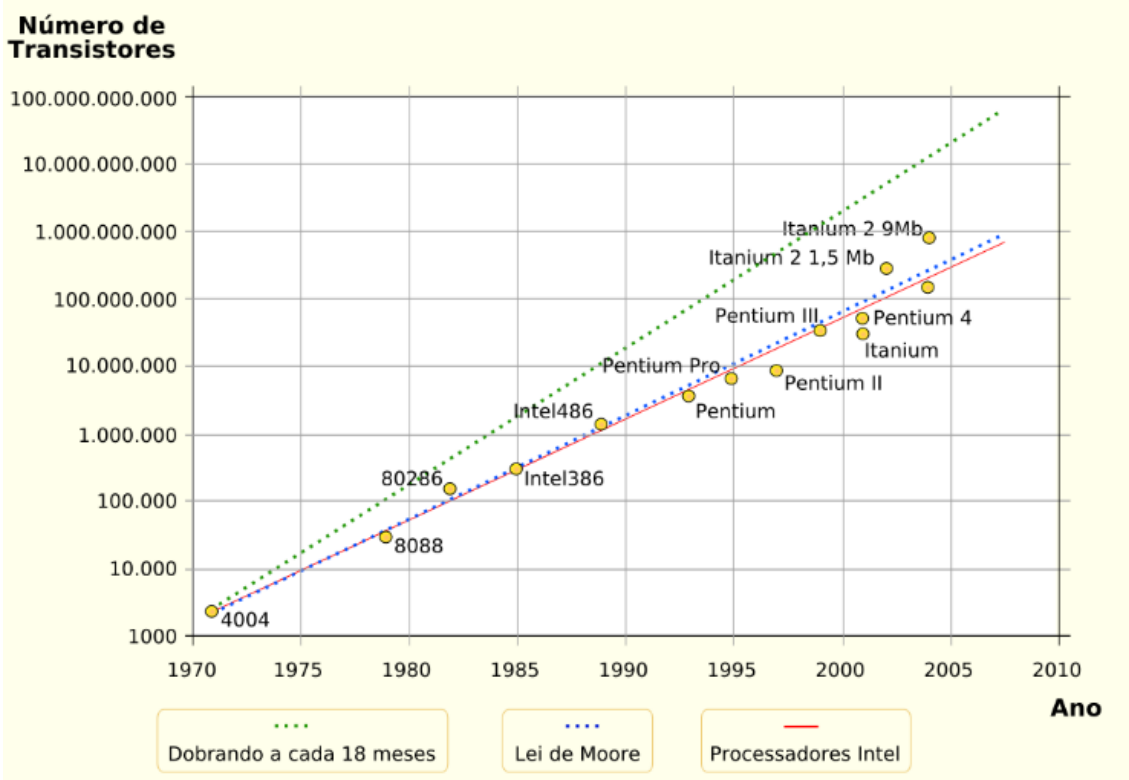


Figura 1. Ley de Moore, evolución de los transistores (gráfico en portugués). *Fuente:* Wikipedia.

Para que podamos entender mejor el impacto de esta ley, se puede afirmar que un teléfono inteligente como el que muchos pueden tener hoy posee la misma capacidad de cálculo que la que tenía la NASA cuando envió al primer hombre a la Luna.

En 2008, un ordenador de 800 euros tenía la misma capacidad de computación que el cerebro de un insecto; en 2013, ya tiene la capacidad de computación de un ratón. Si sigue cumpliéndose la ley de Moore, en 2025 ese ordenador de 800 euros, que estará en el hogar de mucha gente, tendrá la misma capacidad de computación que un cerebro humano. Y en 2045, si sigue cumpliéndose la ley de Moore, ese ordenador tendrá la misma capacidad de computación que todo el conjunto de la humanidad (en ese año, alrededor de 9.000 millones de personas).

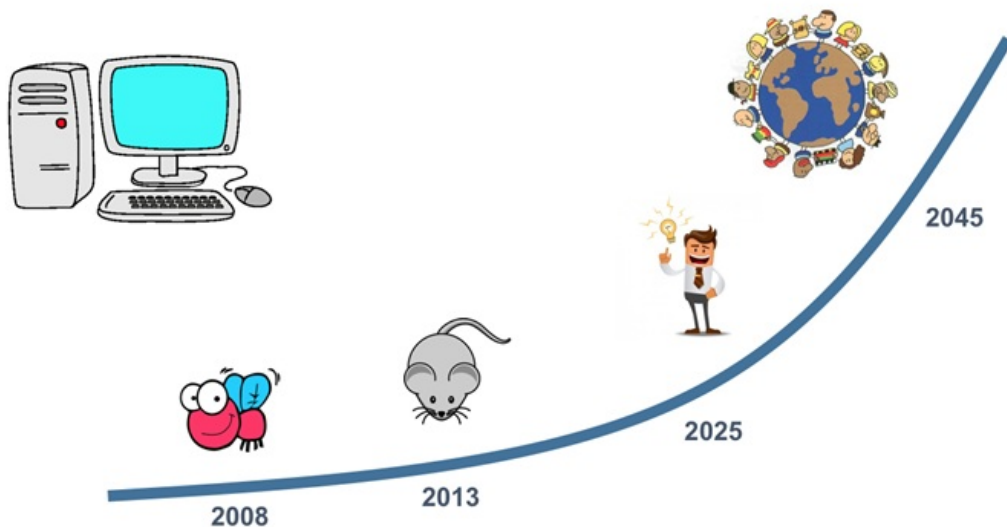


Figura 2. La ley de Moore y su impacto en la capacidad de las máquinas. *Fuente:* elaboración propia.

Por tanto, esta ley pronostica que la capacidad de cómputo de los ordenadores, que era un problema hasta hace pocos años, dejará de serlo y podremos tener una capacidad de cómputo casi ilimitada (esto no quiere decir que sea gratis, pero a un coste determinado tendremos lo que necesitamos).

Durante años, el crecimiento en potencia fue una carrera donde el aumento de potencia venía generado por el incremento en la frecuencia de reloj de las CPU y la disminución en tamaño, que permitía introducir más transistores en los mismos chips. Este crecimiento hacía que la potencia se duplicara casi anualmente, hasta que se llegó al tope de los 3,2-4 GHz. En este punto, el recalentamiento de las CPU hizo imposible continuar evolucionando en esa misma línea.

Fue el momento de incluir los sistemas *multicore*, que continúan hasta día de hoy, donde se combinan CPU con 2, 4, 8 o incluso más CPU de forma coordinada dentro de la misma CPU.

Y más tarde llegó el Big Data, con una aproximación al procesamiento paralelo que volvió a cambiar las reglas del juego.

3.2 Coste de almacenamiento y memoria

Otro elemento clave es el abaratamiento del coste de almacenamiento y de memoria RAM.

El coste de almacenamiento ha bajado de manera exponencial y actualmente está en apenas 0,05 € por giga. Esto, unido a las nuevas tecnologías como las SSD, almacenamiento sólido que permite un acceso muy rápido a la información y cuyo precio se sitúa en torno a los 0,7 € por giga, nos permite almacenar grandes volúmenes de datos a un precio realmente competitivo y, si necesitamos acceder a los datos de manera muy rápida, podemos usar la tecnología SSD.

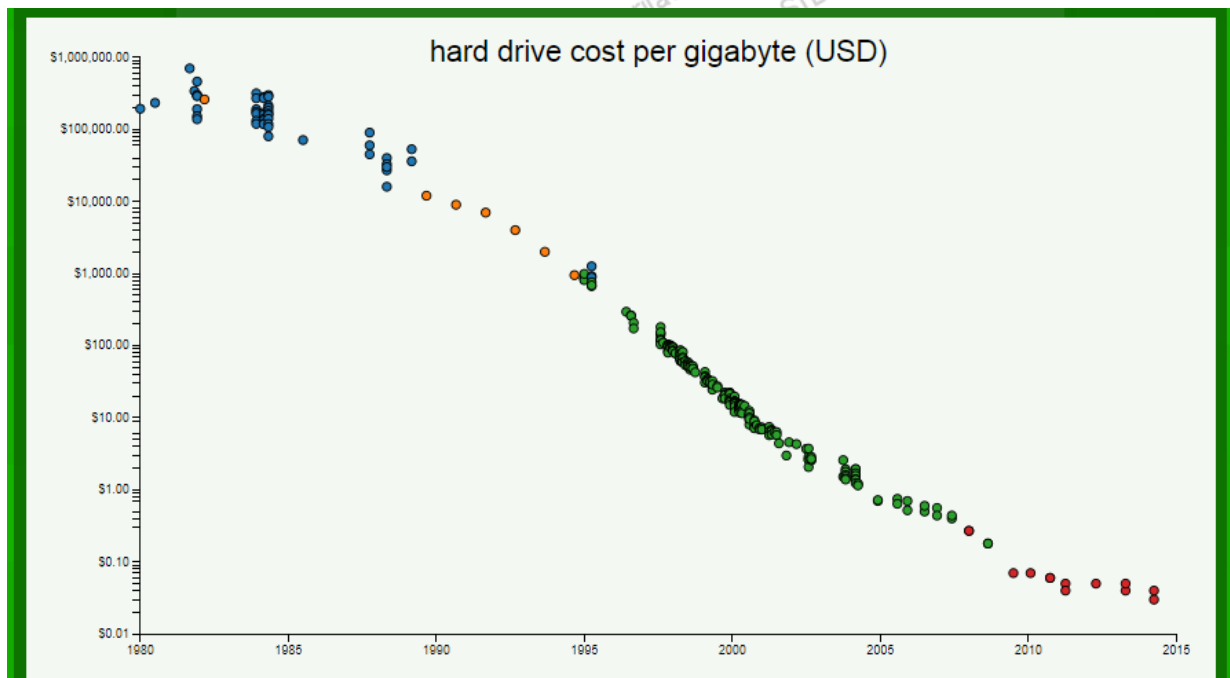


Figura 3. Evolución del coste de almacenamiento. *Fuente:* <http://www.mkomo.com/cost-per-gigabyte-update>

Lo mismo ocurre con la memoria RAM: el incremento de microprocesadores que usan estos dispositivos y los avances en la tecnología han supuesto un incremento de la capacidad de las memorias RAM y una bajada de su precio.

La suma de estos dos factores da como resultado que actualmente podemos procesar cantidades de datos antes impensables, pero no solo eso, sino que esta capacidad de tratamiento de datos está al alcance de cualquier empresa debido a sus reducidos costes.

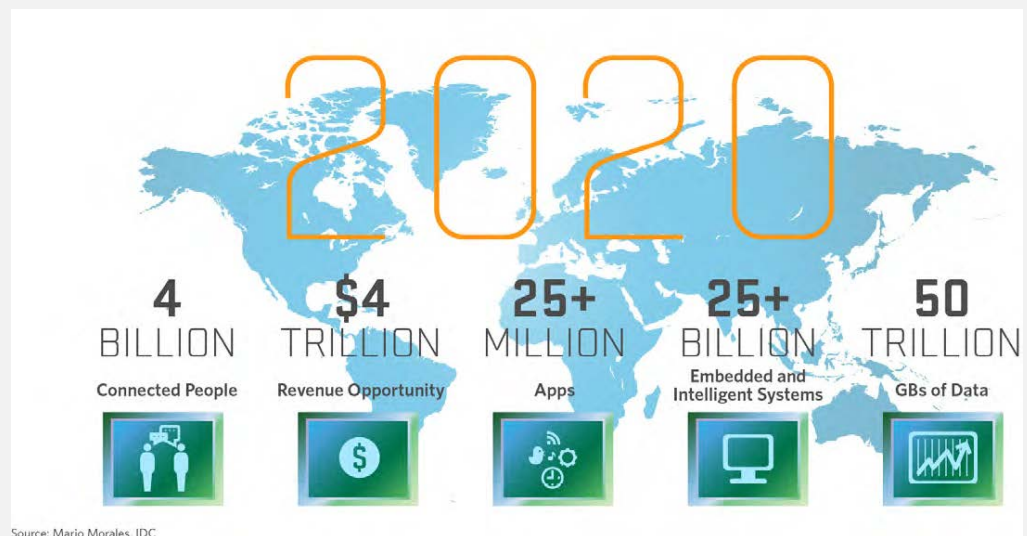
3.3 Digitalización

En una sociedad cada vez más digital, el tiempo que pasamos “conectados” es cada vez mayor. Pero no solo eso, cada vez realizamos un mayor uso de los servicios digitales; es decir, hemos cambiado la forma de interactuar con un gran número de industrias y servicios y ahora consumimos servicios digitales (consultar información, comprar entradas, ver películas, comprar productos, etc.). Como consecuencia, los hábitos del consumidor están cambiando.

Una de las claves del éxito de la digitalización es que los servicios digitales están centrados en el cliente, es decir, dan a cada cliente o usuario lo que espera. Ejemplos de ello pueden ser cuando hacemos una consulta en Google o compramos algo en Amazon; estos nos ofrecen aquello que cada uno de nosotros queremos. Es decir, cuando consultamos en Google, por ejemplo, restaurantes, el buscador no nos devuelve las páginas web que más veces contienen la palabra “restaurante”; en lugar de eso, los resultados muestran los restaurantes que están cerca de nuestra ubicación y, en función de la hora, los más populares para comer o para cenar. En el caso de Amazon, cuando accedemos a su página, nos ofrece de su inmenso catálogo los 20 productos que probablemente más nos interesen y a cada usuario le ofrece productos diferentes en función de sus gustos.

Esto, unido a su conveniencia, es decir, a que su uso es más sencillo para el cliente y así se mantiene o se aumenta en muchos casos la satisfacción de este, explica el éxito de estos servicios digitales y su enorme penetración en el mercado.

El proceso de digitalización es un contexto con el que convivimos hoy en día. Pocos procesos con los que interactuamos de forma diaria son no digitales.



Source: Mario Morales, IDC

Figura 4. 2020 con IoT (*Internet of Things*). Nota: este gráfico utiliza el sistema de numeración anglosajón. Fuente: <https://www.linkedin.com/pulse/2020-internet-things-sunil-aggarwal/>

Como ejemplo, para el año 2020, un estudio de International Data Corporation (IDC) publica que se esperan:

- 4 billones (o 4 mil millones, según el sistema empleado en la mayoría de los países de Europa continental) de personas conectadas, no solamente a nivel de redes sociales, sino a través de algún tipo de dispositivo digital que les permita interactuar, y por tanto que capture información y sea potencial de ser explotado.
- Más de 4 trillones (o 4 mil billones, según el sistema empleado en la mayoría de los países de Europa continental) en oportunidades de negocio dentro del ámbito de la explotación de datos.
- Más de 25 millones de aplicaciones o app, con un incremento del uso de aplicaciones móviles, que van desplazando el uso de aplicaciones web tradicionales.
- Más de 25 billones (o 25 mil millones, según el sistema empleado en la mayoría de los países de Europa continental) de sistemas integrados e inteligentes. Es decir, 25 billones de dispositivos IoT en los cuales podrán integrarse procesos analíticos o recolectar información para explotarla posteriormente.
- Con todo ello, la información que procesar en los próximos años no dejará de crecer. Por tanto, las tecnologías Big Data adquirirán cada vez mayor relevancia como la tecnología óptima para este tipo de procesamiento.

3.4 Efecto red

Vamos a explicar el efecto red con un ejemplo: el teléfono fijo tardó 75 años en tener 50 millones de clientes; la aplicación *Hungry Birds* tardó 35 días. Este efecto también se conoce como "viralidad". Con frecuencia escuchamos que un vídeo o un *post* "se ha hecho viral": eso es el efecto red.

La suma de estos factores hace que los modelos de negocio estén cambiando y que los ciclos económicos se reduzcan drásticamente (inmediatez, agilidad y adaptación al cambio son las claves). Es decir, el tiempo que tienen las empresas para cubrir un servicio nuevo o adaptar sus servicios existentes es muy reducido (alta competencia), por lo que las nuevas organizaciones deben ser ágiles en la reacción al cambio. Como consecuencia del efecto red, el que tiene éxito crece de manera exponencial y se queda o copa gran parte del mercado.

3.5 Aportación del aprendizaje automático

Empresas como Google, Amazon, Netflix, Facebook, Uber, etc., basan sus servicios digitales en tecnologías de analítica avanzada, inteligencia artificial (aprendizaje automático, aprendizaje profundo, redes neuronales) y análisis de redes sociales (teoría de grafos).

La aplicación del aprendizaje automático al mundo empresarial se conoce de manera general como “analítica avanzada”. Como hemos visto, el aprendizaje automático lo podemos definir como el conjunto de técnicas que permiten a las máquinas aprender a resolver un problema basándose en los datos (esta es una de las claves de por qué los datos son un valor estratégico).

En la figura 5 podemos ver una de las claves de por qué la adopción y aplicación del aprendizaje automático es tan amplia y disruptiva.

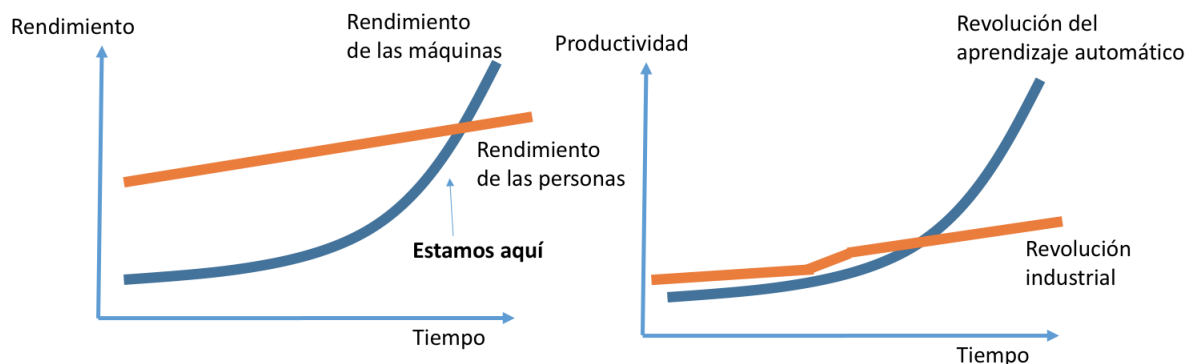


Figura 5. Capacidad de las personas comparada con la evolución de la capacidad de las máquinas. Productividad obtenida con la revolución industrial comparada con la productividad obtenida con aprendizaje automático. *Fuente:* elaboración propia.

La productividad que conseguimos cuando enseñamos a las máquinas a hacer determinadas tareas es exponencial; en la figura 5 podemos ver una comparación temporal entre la productividad de la revolución industrial y la productividad cuando se aplica el aprendizaje automático.

Si bien la capacidad de cómputo, incluso con la ley de Moore, era un cuello de botella, el desarrollo de las tecnologías GPU (unidades gráficas de procesamiento) ha resuelto este problema. Como dato, en capacidad de cómputo, una GPU puede equivaler a más de 100 ordenadores.

Para evaluar el impacto de todo lo que hemos comentado, podemos decir que antes de esta revolución la empresa que tenía el producto tenía el mercado; sin embargo, actualmente es la empresa que proporciona el mejor servicio al cliente la que ocupa el mercado.

Facebook, la red social más popular del mundo, no genera contenidos; Uber, la compañía más grande de transporte de personas, no tiene taxis (si excluimos sus servicios *black*); Alibaba no tiene inventario y Airbnb no tiene alojamientos propios.

En la figura 6 vemos la evolución bursátil de las compañías de referencia en este ámbito; podemos observar que su capitalización no para de subir.



Figura 6. Evolución de la capitalización bursátil (en verde, las compañías que aplican aprendizaje automático).

Fuente: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation.aspx?linkId=38285932>

Aquellas compañías que han apostado por la aplicación sistemática de estas técnicas de analítica avanzada han experimentado una fuerte revalorización bursátil, reflejo directo de sus beneficios.

IV. El valor del dato

Los datos hay que transformarlos para darles sentido de negocio. Es decir, un dato no elaborado no aporta información.

En la distribución de la pirámide informacional, vemos que los datos crudos deben transformarse en información, la información en conocimiento y el conocimiento en inteligencia.

La forma de pirámide no solo representa las fases; pretende indicar que, para obtener conocimiento y después inteligencia, necesitamos muchos datos crudos o no elaborados. Es importante destacar esto para entender mejor por qué necesitamos tecnologías específicas para el tratamiento de grandes volúmenes de datos.

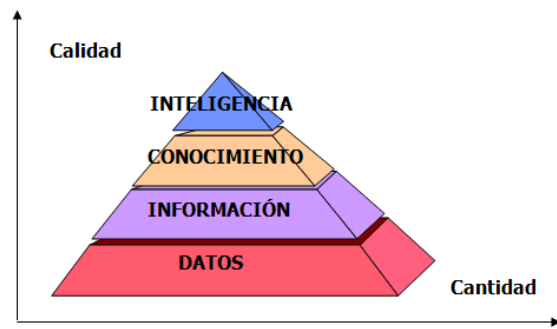


Figura 7. La pirámide informacional. *Fuente:* Ponjuán (1998).

En esta pirámide, la última fase es la inteligencia, es decir, la capacidad de tomar decisiones basándonos en los datos. Esta es una de las claves del Big Data.

Pensemos un momento en los datos. Los datos que tiene una compañía son consecuencia de su actividad, de los productos o servicios que comercializa, de la política de precios, de la política de recursos humanos, de los propios recursos humanos (empleados), de los procesos internos y de su forma de interactuar con el cliente. Es decir, los datos son en esencia la consecuencia de la actividad de una compañía; por eso aprender de los datos es clave y los datos son un valor estratégico.

4.1 Impacto del dato (Big Data y Analytics) en las organizaciones

En la actualidad, el dato como activo de una empresa es una realidad. Es decir, usamos la información disponible como un activo más de la empresa para generar nuevas ventajas competitivas y aquellas empresas que no utilizan el dato de esta manera están en clara desventaja competitiva.

Pero usar los datos o la información como un activo debe tener un objetivo claro. Por ello, la primera pregunta que debemos hacernos es: ¿qué beneficios aportan las tecnologías Big Data y Analytics mediante el uso de datos a las organizaciones?

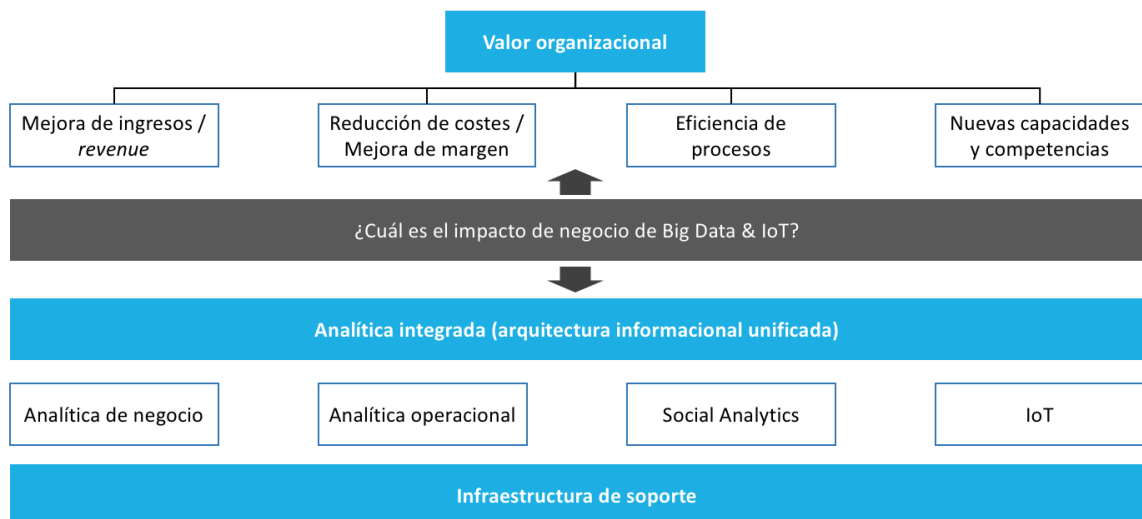


Figura 8. Aportación de valor de Big Data y Analytics. *Fuente:* elaboración propia.

A continuación, veremos a qué se refiere cada uno de estos aspectos.

4.1.1. Mejora de ingresos / *revenue*

Se refiere a la mejora en penetración de mercado y penetración en nuevos mercados. En este apartado se engloban todas las iniciativas que permiten obtener mayores ingresos de un cliente, obtener más y mejores clientes o incluso analizar nuevos mercados donde poder posicionar los servicios de una empresa. Por ejemplo:

- Mejorar ventas creando mejores perfiles de cliente, atracción de nuevos clientes y venta *cross-selling* y *up-selling*.
- Capacidad para crear nuevos mercados o proporcionar nuevos servicios sobre la cadena de valor del cliente.

4.1.2. Mejora de márgenes

Referido a la reducción de costes directos e indirectos en bienes y servicios y mejor retorno de inversión. En este caso, reducción de costes por el uso de tecnología Big Data (reducción del coste de cómputo y de almacenamiento del dato) y procesos analíticos que permitan reducir el coste de la operación de negocio del cliente, como:

- Reducir gastos en la cadena de valor y mejorar la agilidad para responder a picos de demanda o cambios de comportamiento del cliente.
- Mejor retención del cliente y grado de satisfacción con mejores estrategias de precio y gestión de la entrega, por ejemplo.

4.1.3. Mejora de eficiencia

Mejora de la eficiencia entendida como un mejor equipamiento y mantenimiento de los sistemas, tanto en su coste como en su disponibilidad. Aquí no nos referimos a mejoras por cambios de tecnología sino al uso de la información disponible de los sistemas para incrementar su eficiencia, como por ejemplo:

- Duración mayor de equipamiento y mejora de la eficiencia del equipamiento actual.
- Mejor uso a través de una gestión óptima de la demanda.

4.1.4. Competencia

Se refiere a la mejora de la percepción de mercado, mejorar la capacidad de responder a factores externos y reducir el *time-to-market* de nuevos servicios o productos. A su vez, en este apartado también se incluyen nuevos productos o servicios que se generan mediante el uso de información. Como ejemplos:

- Mejora del precio de *stock* y valor de la compañía, mejora de relaciones y una respuesta más rápida a amenazas externas.
- *Time-to-market* significativamente menor.

4.2 Áreas de gestión del dato

De cara a obtener los beneficios vistos en el apartado anterior, es necesario contar con un proceso completo del dato que permita desde su captura y gestión hasta su posterior explotación y operativización. Para ello es necesario conocer la cadena de valor del dato, ya que las diferentes iniciativas que se generen dentro de los *Business Case* deberán contar con elementos de toda esa cadena de valor o bien requerir que se desarrolle para poder ejecutarlos.

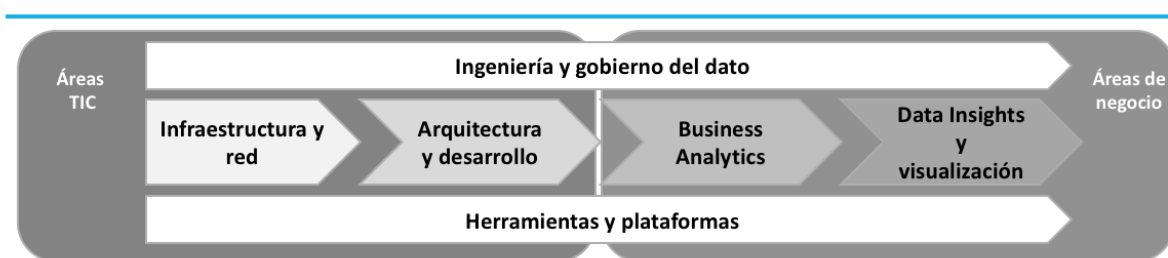


Figura 9. Cadena de valor del dato. *Fuente:* elaboración propia.

Como puede verse en la figura 9, la cadena de valor del dato tiene sistemas más enfocados a las áreas TIC o departamento de IT, mientras que otra parte está más enfocada en las áreas de negocio.

Ingeniería y gobierno del dato

Se refiere a los procesos necesarios para asegurar que se gestiona el gobierno de la información. En este apartado se incluyen los procesos de:

- Calidad del dato.
- Linaje y trazabilidad.
- Análisis de impacto.
- Diccionario semántico.
- Otros.

Infraestructura y red

Se refiere a la infraestructura necesaria para la captura, almacenamiento y explotación de la información. Por tanto, se refiere a los sistemas físicos, tanto *on-premise* como *cloud*, necesarios para dotar a la organización de las capacidades analíticas y de gestión del dato que necesitan. Habitualmente se requiere analizar:

- Capacidad de almacenamiento, separando por el ciclo de vida del dato (datos calientes o de alta carga de acceso, datos fríos o de baja carga de acceso y almacenamiento para cumplimiento normativo o procesos puntuales de baja periodicidad). Esto afectará a la calidad de los sistemas de almacenamiento necesarios y a su coste.
- Capacidad de procesamiento, entendida como la potencia de cálculo disponible para procesos de tratamiento y gestión del dato, así como para procesos analíticos.

Arquitectura y desarrollo

Se refiere a los procesos que se encargan de transformar la información desde los formatos de origen hasta un formato óptimo para su posterior explotación.

Business Analytics

Procesos de analítica y *machine learning* aplicables a los datos para generar los resultados predictivos o prescriptivos necesarios.

Data Insights y visualización

Procesos para comunicar los resultados de Business Analytics a las áreas de negocio y operativizar los modelos, es decir, cómo las áreas de negocio van a utilizar los resultados de los modelos matemáticos en su operativa diaria.

Herramientas y plataformas

Se refiere a las herramientas y sistemas necesarios para sustentar el resto de los procesos presentados.

V. Tipos de datos desde la perspectiva de negocio

Nos encontramos ante una auténtica explosión de datos. Las empresas están captando cada vez más datos de las transacciones con sus clientes, las interacciones en redes sociales o los sensores diseminados en el entorno.

A mediados de la primera década del siglo XXI, IDC estimaba el volumen total de datos en el mundo en alrededor de 130 exabytes (1 exabyte son 1.000 millones de gigabytes). En 2010, esta cifra se había multiplicado por 10, y en 2013 se estima que llegó a los 8.000 exabytes (alrededor de 60 veces más). Por ejemplo, Twitter procesa aproximadamente 500 millones de tuits cada día. Cada minuto se suben a YouTube más de 100 horas de vídeo. Netflix envía a sus suscriptores más de 2.400 millones de horas de vídeo cada mes y la base de datos de Facebook supera los 100 petabytes.

Debemos hacer una primera clasificación de los datos en cuanto a su aplicación empresarial:

Datos disponibles

Todos los datos con los que cuenta una empresa.

Datos inteligentes

La fracción de los datos disponibles que son de utilidad para solucionar un problema o proveer un servicio.

Datos enriquecidos

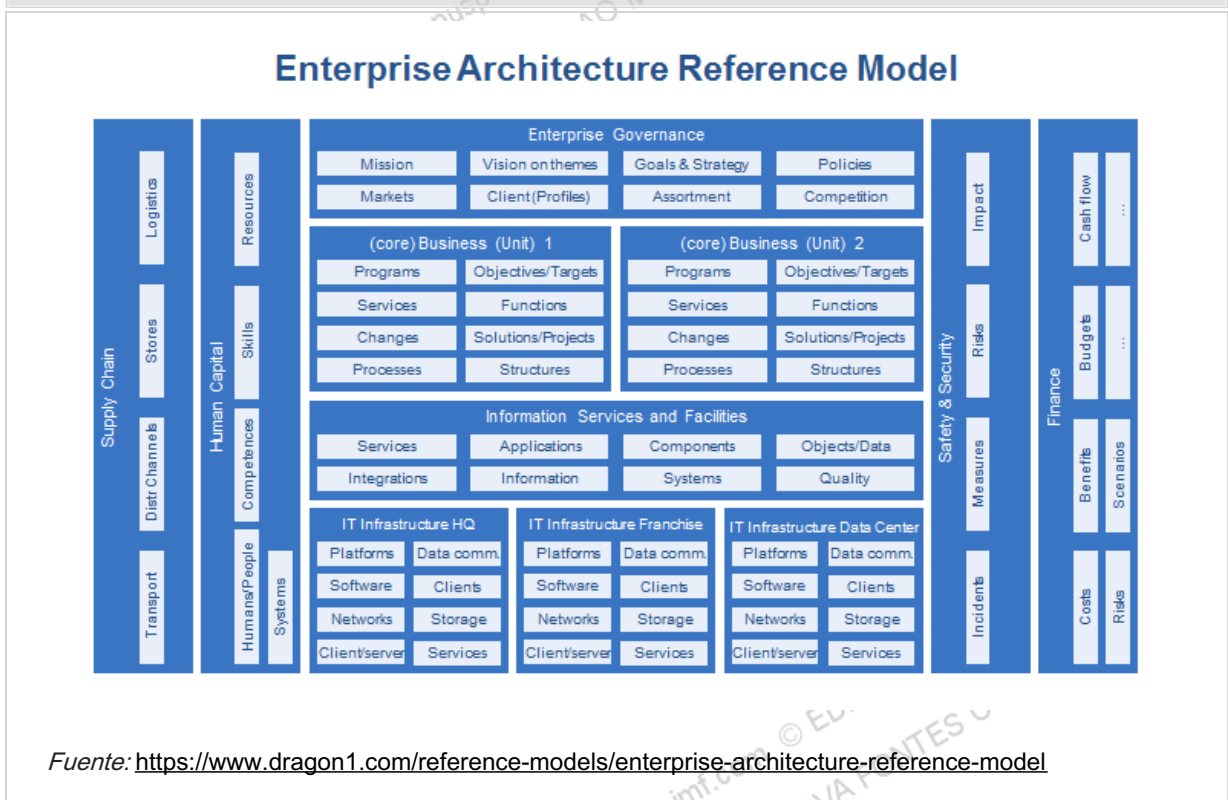
Aquellos datos inteligentes que se enriquecen con otros datos adquiridos de fuentes externas. Por ejemplo, si queremos optimizar el sistema de recogida de basuras, podemos obtener de internet información sobre las casas en venta en cada zona, ya que es de esperar que haya reformas; si queremos conocer la aceptación de un servicio, o si queremos anticiparnos a la no satisfacción de un servicio para tomar medidas antes de que sea un problema, podemos atender a las redes sociales. Esta es una de las claves: encontrar datos que enriquezcan los que tiene la compañía para poder tomar mejores decisiones.

Las empresas tradicionales están usando aproximadamente entre un 7 y un 12 % de la información que tienen para la toma de decisiones y para dar servicio a sus clientes, mientras que las empresas digitales nativas usan aproximadamente un 70 %. Las empresas tradicionales se han dado cuenta de que tienen muchos datos que actualmente no están usando, pero que tienen un gran potencial en cuanto a impacto de negocio se refiere. **Los datos son el petróleo del siglo XXI.**

Gran parte de la limitación del uso de la información viene dada por la arquitectura de sistemas tradicional, en la que las diferentes áreas de negocio y las diferentes funciones dentro de cada una mantienen sus propios sistemas de información y crean silos que dificultan el análisis completo de los problemas.

En la figura 10 se expone un modelo de referencia de arquitectura de sistemas de una gran empresa. Podemos ver que el modelo de referencia crea silos de información, lo que hace muy complejo el análisis completo de un proceso, es decir, la información de logística, la de operaciones, financiera, recursos humanos, etc.

Figura 10. Marco de referencia para la arquitectura de empresas.



Las empresas se enfrentan al reto de procesar los datos con el objetivo de transformarlos en conocimiento y en servicios.



Para ello, encuentran varios desafíos:

- Ser capaces de almacenar los datos de manera que se puedan analizar procesos completos y no solo partes aisladas.
- Ser capaces de integrar los datos que vienen de numerosas y diversas fuentes. Es decir, conectar los datos que vienen de diferentes silos de información dentro de la empresa y enriquecerlos con datos adicionales externos.
- Ser capaces de procesar los datos y transformarlos.
- Ser capaces de analizar los datos para identificar los que son relevantes y tienen valor de negocio.
- Ser capaces de buscar fuentes externas para enriquecer los datos.
- Transformar los datos en conocimiento, es decir, en información relevante para el negocio o servicio.
- Desarrollar o hacer evolucionar los procesos actuales a sistemas o aplicaciones basadas en los datos que proporcionen valor, bien para la toma de decisiones, bien para el desarrollo de servicios digitales (centrados siempre en el cliente).

Los avances de las tecnologías de procesamiento masivo de datos, así como la aplicación de las nuevas técnicas analíticas (aprendizaje automático), proporcionan a las empresas la capacidad de hacer cada vez más cosas con esos datos y afrontar los retos que hemos comentado, utilizándolos no solo para entender el pasado, sino para predecir el futuro, tomar decisiones de manera automatizada y centrar su actividad en el cliente. De este modo, las compañías están empezando a utilizar los datos para mejorar sus procesos de negocio, incrementar su productividad, hacer más eficientes sus operaciones, entrar en nuevos mercados o reinventar sus modelos de negocio.

VI. Tipos de analítica

El Big Data representa un cambio de paradigma en la gestión y la forma de hacer negocios. Ahora, con el Big Data y la analítica avanzada, **preguntamos a los datos qué es lo que va a pasar en lugar de qué ha pasado (BI)**. Por ejemplo: cuánto voy a vender, cuánta gente vendrá a mi tienda, cuál es la mejor ruta para ir de un punto a otro, qué producto le va a interesar a un cliente, etc.

Poder explotar los datos y ser una compañía que base sus decisiones en datos no es algo sencillo. Debemos tener en mente los diferentes tipos de analítica que existen y entender que no podemos saltar al último escalón sin haber pasado por los anteriores.

Tipos analítica en la toma de decisiones

Existen cuatro tipos de analítica que nos llevan a diferentes estadios en la toma de decisiones. Se enumeran en orden de dificultad:

Analítica descriptiva

Se analizan los datos aplicando diferentes técnicas estadísticas para entender qué es lo que ha ocurrido.

Analítica de diagnóstico

Es el siguiente paso y aplicamos técnicas de causa-efecto para entender el motivo por el que ha ocurrido un hecho.

Analítica predictiva

Aplicando técnicas de aprendizaje automático e inferencia estadística, se intenta responder a qué es lo que va a ocurrir. Da respuesta a una pregunta basándose en los datos.

Analítica prescriptiva

Se basa fundamentalmente en una técnica de aprendizaje profundo que se conoce como aprendizaje por refuerzo e intenta decirnos qué debemos hacer para que algo ocurra.

Los diferentes tipos de analítica aportan distintos valores en la cadena estratégica, pero la que mayor valor aporta es la analítica prescriptiva, aquella que nos dice qué es lo que debemos hacer para conseguir un objetivo.

Tipos de analítica en función de áreas de negocio o impacto

Dentro de la analítica, a su vez, existe una distinción por áreas de negocio o impacto:

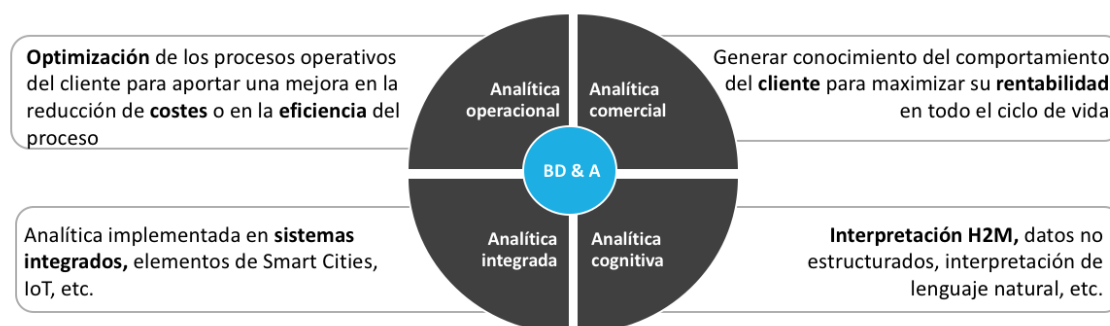


Figura 11. Tipos de analítica en función de áreas de negocio o impacto. *Fuente:* elaboración propia.

Analítica operacional

Enfocada a la optimización de los procesos internos de una compañía y con un claro énfasis en la reducción de costes y mejora de la eficiencia de los procesos.

Analítica comercial

Se trata de todos aquellos procesos analíticos enfocados a mejorar el conocimiento del cliente y ayudar en el soporte a la toma de decisiones tanto en ámbitos comerciales como de marketing; retención de clientes, etc. Su objetivo principal es la mejora de ingresos e incremento de la cuota de mercado, a la vez que se mantiene la satisfacción del cliente y se evita su fuga.

Analítica integrada

Se trata habitualmente de los procesos analíticos sobre sensores o sistemas integrados, elementos principales del IoT, Smart Cities, etc. Su objetivo, por tanto, es ayudar en procesos dentro de elementos integrados, como el vehículo autónomo.

Analítica cognitiva

Se trata de la interpretación humano-máquina y viceversa. Habitualmente se trata de procesos de tratamiento de datos no estructurados que conllevan la interpretación de imágenes, lenguaje natural, etc.

Tradicionalmente, la mayoría de las decisiones empresariales se tomaban en función de la experiencia en entornos con poca información. El análisis avanzado de datos permite obtener información objetiva sobre el pasado e incluso predicciones sobre el futuro, a partir de las cuales tomar mejores decisiones, basadas ya no en la experiencia sino también en datos, en hechos (aquello que ha ocurrido).

Una compañía debe adquirir competencias de analítica predictiva y prescriptiva para poder tomar decisiones basadas en datos. Es necesario aprender a trabajar en este nuevo contexto, comprender que muchas decisiones que antes se tomaban de manera intuitiva, ahora vendrán inequívocamente determinadas por el análisis de la información disponible.

En este punto, podemos preguntarnos qué valor aporta a una empresa transformarse en una compañía que toma decisiones basadas en datos.



Anotación: Proporciona varios beneficios

- **Productividad.** Hemos visto que la aplicación de las técnicas de analítica avanzada proporciona gran productividad.
- **Agilidad.** La capacidad de las máquinas es muy superior a la de las personas; los algoritmos pueden reentrenarse (o incorporar nuevas variables) cuando pierden eficacia y adaptarse así a los nuevos escenarios.
- **Objetividad.** Nos basamos en hechos para tomar decisiones, no en intuiciones.
- **Centrado en el cliente.** Podemos desarrollar servicios que nos digan que es lo que cliente quiere con mayor probabilidad.

VII. Puntos clave en la transformación

La analítica avanzada y el Big Data suponen un cambio cultural, tecnológico y organizativo que sitúa a las empresas ante numerosos retos. Las empresas tienen que enfrentarse a estos retos acometiendo importantes transformaciones en su forma de trabajar y generar valor (deben centrarse en el cliente).

Con el fin de volver a tener el control del proceso de compra, desde la búsqueda del producto a la compra final y posterior fidelización, vendedores y proveedores de servicios han desarrollado capacidades analíticas y de tratamiento de datos para comprender a sus clientes, tratar de anticipar sus próximos movimientos y posicionarse en su proceso de decisión de compra.

La mayoría de las organizaciones reconocen que la adopción del Big Data es una parte esencial del camino hacia la transformación digital. La experiencia digital del consumidor pasa por conocerlo y entenderlo, y eso significa recoger, analizar e interpretar los datos obtenidos de la experiencia del cliente desde todos los dispositivos y puntos de contacto con él, ya sean internos o externos, como las redes sociales. Ese conocimiento del cliente permitirá desarrollar nuevos productos o servicios con un alto grado de personalización y más rentables².

Para que podamos entender la transformación profunda a la que nos enfrentamos, es interesante leer un tuit de Andrew Ng, una de las personas más influyentes dentro de la aplicación de las técnicas de analítica avanzada al mundo empresarial, uno de los padres de Google Brain y uno de los impulsores clave del coche autónomo: "Muchos CEO de las compañías de S&P 500 desearían tener una estrategia de internet. En 5 años oiremos lo mismo referido a la estrategia de inteligencia artificial"³. Se puede entender que la analítica avanzada (inteligencia artificial) va a ser clave y estará liderada desde la alta dirección, es decir, que es estratégica.

Es evidente que esta transformación debe estar liderada por la alta dirección, y requiere:

Nuevo liderazgo

La gestión empresarial requiere visión. A través de ella, se pueden detectar oportunidades y se puede entender cómo evolucionan los mercados. Solo esta visión hace posible pensar creativamente e innovar. Las tecnologías analíticas proporcionan a los líderes una visión única de la realidad, antes no disponible para las empresas. Los líderes son los responsables de generar una visión empresarial y de inspirar a los equipos para alcanzar objetivos. Las empresas necesitan nuevos líderes que sean capaces de trabajar en el nuevo contexto empresarial de los datos y proporcionar sentido al conocimiento extraído de su análisis.

Tecnología

Dado el volumen y la variedad de los datos generados hoy en día, las compañías necesitarán invertir en tecnología que les permita captar, analizar y presentar esos datos de una forma eficiente. Sin embargo, la tecnología más innovadora es estéril sin una estrategia analítica que permita realimentar los procesos de negocio para optimizar su diseño y rendimiento. En la actualidad, las empresas siguen fundamentalmente invirtiendo en tecnologías para *explicar el pasado* y descubren tendencias subyacentes en los datos difícilmente descubribles sin un uso intenso de la tecnología. Menos universales son las tecnologías para *predecir* comportamientos futuros y *actuar* a la luz de ese conocimiento. La buena noticia es que empiezan a estar disponibles tecnologías viables para ello, por lo que la tecnología no será el principal freno a la incorporación de la analítica avanzada en las empresas.

Gestión del talento

En la economía de los datos, las empresas necesitarán disponer de talento analítico. Entendemos por talento analítico el mostrado por profesionales capaces de extraer valor de los datos poniendo en contexto los resultados y derivando los aprendizajes que la empresa debe extraer de ellos, así como las decisiones que dicho conocimiento pueda provocar. Este talento puede residir en las áreas funcionales (como profesionales de negocio con conocimientos analíticos), o en el departamento TIC, con especialistas en datos. Los conocimientos que se necesitan, además de los tecnológicos, caen en los territorios de las matemáticas y, especialmente, de la estadística, claves para la analítica predictiva; sus competencias deben incluir la de comunicar de manera significativa los resultados obtenidos, el pensamiento conceptual o la capacidad analítica.

Cultura empresarial

Es necesario que se produzca un cambio cultural en las organizaciones para adaptarse al nuevo contexto del mundo de los datos. Las organizaciones centradas en los datos fomentan la toma de decisiones informadas, el análisis en tiempo real y la minimización de las barreras jerárquicas y organizativas. Los integrantes de la organización han de compartir este punto de vista y aplicarlo a su trabajo diario para mejorar su desempeño y el de todos y poder experimentar que las decisiones basadas en datos tienden a ser mejores decisiones.

²“El Big Data y Analytics, elementos clave de la transformación digital”, Los blogs de *La Razón*. [En línea] URL disponible en: <https://www.larazon.es/blogs/tecnologia-redes-sociales/transformacion-digital/big-data-y-analytics-facilitadores-de-la-transformacion-digital-GB11017203>

³*Tweet* de Andrew Ng, Twitter. [En línea] URL disponible en: <https://twitter.com/andrewyng/status/785478101912256512>

VIII. La organización analítica

Las empresas están actualmente buscando o definiendo su estrategia de transformación digital y transformando sus estructuras. En los próximos años veremos como las empresas y los comités de dirección comienzan a definir la estrategia corporativa de analítica avanzada (inteligencia artificial); una forma de entender el Big Data y la inteligencia artificial es la transformación de la transformación digital, de ahí su gran impacto en todos los ámbitos y su carácter disruptivo.

Estos cambios suponen un reto para las compañías que deben adaptarse a esta nueva realidad económica. Podemos agrupar los retos que afrontan las organizaciones en dos grandes ámbitos:

Organizativo

Las compañías deben adaptar su organización y estrategia para poder competir en un entorno centrado en el cliente y cada vez más digital. El objetivo de toda compañía debe ser convertirse en una empresa que toma decisiones basándose en datos y aplica el aprendizaje automático de manera sistemática.

Talento

Como consecuencia de esta transformación, debemos incorporar o formar empleados para que tengan las competencias requeridas por los nuevos perfiles, tanto directivos como tecnológicos. El cambio en la organización debe ir acompañado de un cambio en la cultura laboral de la empresa. Este cambio cultural debe estar liderado por directivos con visión estratégica, con grandes habilidades de liderazgo y con experiencia en gestión del cambio y transformación.

Vamos a comentar las diferentes opciones para organizar el área analítica. Las iremos enumerando de menor a mayor grado de madurez de la analítica en la organización.

Concentración en un área funcional

El equipo analítico reside en una única área funcional, probablemente la que tiene una visión más clara de la necesidad y utilidad de las iniciativas analíticas, en general finanzas o marketing. Este modelo se ve impulsado normalmente por las áreas funcionales. Tiene varios inconvenientes: existencia de silos, tanto de conocimiento como de datos, e inexistencia de cultura analítica. En muchas ocasiones es la forma en que muchas empresas comienzan con la analítica, centrándose en el área de mayor necesidad. Una vez se ha demostrado el valor y tenemos procesos, podemos pasar al siguiente modelo, el centralizado, donde la analítica del área funcional pasa a ser un área corporativa.

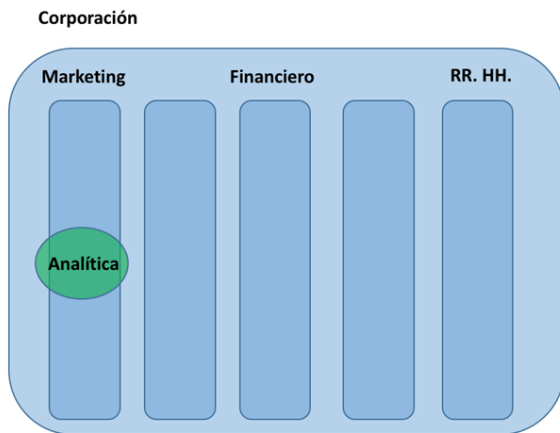


Figura 12. Organización enfocada en un área funcional. *Fuente:* elaboración propia

Modelo centralizado

Se crea un equipo en una única ubicación geográfica y organizativa (o se hace corporativa la analítica del área funcional). Este equipo es el que gobierna la estrategia analítica y el que despliega a los analistas para cada una de las áreas de negocio en las que son necesarios. En este modelo suele existir un “responsable de la analítica” y un equipo de análisis funcional que da soporte a todas las unidades de negocio. Es un buen modelo para aprovechar las sinergias, tener un buen modelo de datos y asegurar la calidad de estos. Es el modelo adecuado cuando estamos comenzando, ya que tenemos que concentrar el escaso talento del que disponemos.

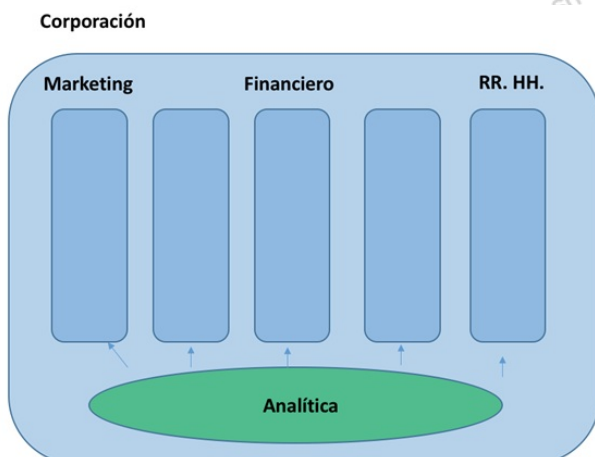


Figura 13. Modelo funcional. *Fuente:* elaboración propia.

Centro de excelencia o competencia

En este modelo, existe un equipo centralizado que da soporte a las áreas funcionales, pero estas tienen analistas propios, a los que sirve el centro de excelencia para proporcionarles consejo, tecnología, modelos y mejores prácticas. Es excelente para la implicación de las áreas funcionales y el alineamiento de las iniciativas con las necesidades del negocio. Es decir, comenzamos a expandir el conocimiento y aprovechamos a los analistas de las áreas funcionales para primero darnos soporte y después irlos formando para que adquieran las capacidades analíticas necesarias. Es el centro de excelencia el que define las políticas corporativas, la organización de datos, técnicas y herramientas, la metodología y el que se encarga de compartir el conocimiento o asegurarse de que se comparte el conocimiento entre los diferentes departamentos.

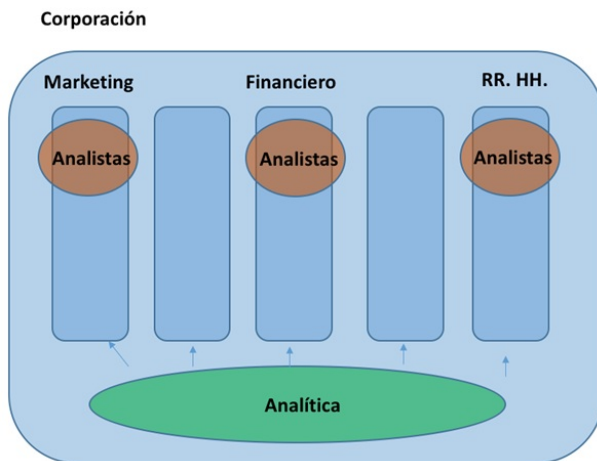


Figura 14. Organización basada en centro de excelencia o competencia. *Fuente:* elaboración propia.

Descentralizado

Cada área funcional o geográfica tiene su propio equipo y herramientas analíticas. Proporciona una máxima agilidad a las áreas, pues adapta las inversiones a las necesidades reales. La organización descentralizada, siempre con un director de datos y analítica que aporte las líneas maestras que debe seguir cada área funcional, es la más productiva ya que tenemos expertos en cada área.

Se comete un error muy común en las organizaciones, que es intentar saltar al modelo descentralizado de manera anticipada o como primer modelo de organización. En ese caso corremos el riesgo de enfrentarnos a los siguientes problemas:

- ➔ No hay unas herramientas, método y puesta en común del conocimiento. El mismo error se puede dar en varias áreas, no se comparten ni éxitos ni lecciones aprendidas.
- ➔ No tenemos una política de datos y analítica común.
- ➔ No existe estrategia de datos.

Corporación

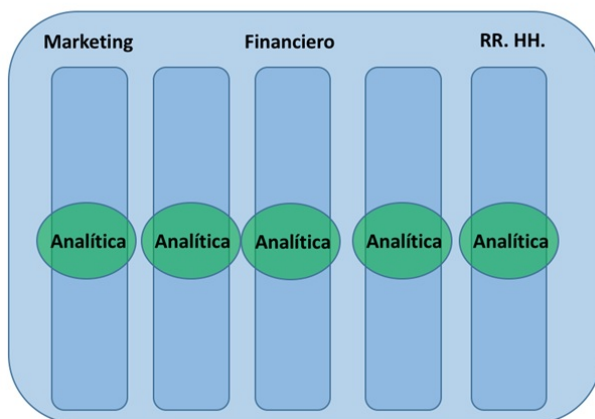


Figura 15. Modelo descentralizado. Fuente: elaboración propia.

IX. El Big Data y la empresa

Hasta el momento, hemos analizado el contexto económico, el valor de los datos en la organización y cómo este contexto está transformando las organizaciones empresariales para poder convertirse en compañías que basan sus decisiones en datos y usar de manera masiva el aprendizaje automático.

Esto justifica solo en parte el auge de las tecnologías de uso masivo de datos como puedan ser Hadoop, HDFS, Spark, Flink y otras. En este epígrafe nos detendremos en los motivos de este auge.

9.1 Costes

Uno de los motivos clave de este auge es el coste: las tecnologías Big Data son en general *open source* y cualquiera puede bajarse las distribuciones de Apache y construir un clúster. Evidentemente, las organizaciones no suelen recurrir a esta forma de usar el Big Data, se apoyan en proveedores que les garantizan que todas las herramientas del ecosistema están configuradas correctamente y proporcionan también soporte técnico.

Aun con estos costes, las soluciones Big Data son mucho más competitivas, aunque según la solución por la que optemos (si es en la nube o con una instalación en nuestro centro de datos) los ahorros de costes pueden estar entre el 20 y el 50 % en función del proveedor.

Cualquier compañía, a la hora de tomar una decisión de inversión, debe hacer un balance de los costes que deberá soportar. Por ello, siempre que se trata el impacto de la implementación del Big Data sobre los costes se pone el énfasis en el almacenamiento de datos, en la cantidad de megas, teras y zetas que se necesitan, en cómo combinar un almacenamiento de disco duro tradicional y flash para contener los costes manteniendo la agilidad en el acceso a esos datos, etc. Pero existen otros aspectos importantes que deben tenerse en cuenta a la hora de valorar dar el paso y realizar esta inversión. La implantación del Big Data tiene su ventaja más notable en la drástica reducción en costes de tiempo. Para un 97 % de los profesionales del Business Intelligence, los procesos de extracción, transformación y carga son críticos y suponen de un 50 a un 90 % de sus horas de trabajo⁴.

Esta mejora en la rapidez para la obtención de información relevante para la toma de decisiones permite a la organización adelantarse en su estrategia con respecto a sus competidores y beneficiarse de ser la primera⁵.

⁴Alexandre Peretó. "El impacto del Big Data sobre los costes". Blogs ICEMD, blog La gestión del Big Data en la inteligencia de negocio. ESIC, 26 de octubre de 2015. [En línea] URL disponible en: <http://blogs.icemd.com/blog-la-gestion-del-big-data-en-la-inteligencia-de-negocio-el-impacto-del-big-data-sobre-los-costes/>

⁵Ibidem.

9.2 Agilidad, flexibilidad

Otro aspecto muy importante es la capacidad para poder tratar no solo una cantidad de datos muy alta (terabytes o petabytes), sino para tratar datos de cualquier naturaleza usando la misma plataforma. De esta manera, incluir nuevas fuentes de datos es solo cuestión de añadir la nueva fuente de datos y definir su procesamiento.

También nos proporciona la opción de tratamiento por lotes o en tiempo real usando la misma infraestructura, lo que nos permite tener una infraestructura común para todas las aplicaciones de la compañía y reducir los costes de administración al centrarnos en una única plataforma.

Actualmente existe una tendencia a incorporar GPU como elementos de procesamiento de alta velocidad.

La naturaleza de procesamiento de las GPU aumenta drásticamente el paralelismo de la computación en el sistema y redundando en un aumento de la potencia de procesamiento, lo que disminuye los tiempos necesarios para procesar los datos, el coste de los cálculos y, sobre todo, permite resolver problemas imposibles de escalar utilizando CPU. Esto es especialmente interesante en la construcción de sistemas basados en *machine learning*, en los que un mayor número de *cores* de computación mejoran notablemente los tiempos puesto que este tipo de soluciones resuelven una enorme cantidad de operaciones matemáticas que pueden realizarse a la vez y, por tanto, aprovechan toda la paralelización de una GPU y una enorme cantidad de *cores*. Gracias al incremento en la capacidad de cómputo, en entornos de ciencia de datos, el uso de una GPU acorta drásticamente el tiempo empleado en los ciclos de análisis, entrenamiento de modelos y despliegue en producción; y permite al científico de datos probar las hipótesis en tiempos menores y mejorar el entendimiento de los datos y la precisión. Procesado de textos, sistemas de recomendación, computación visual... son algunos de los campos en los que destaca la aplicación de tecnologías GPU por su reducción de tiempos y costes⁶.

⁶BEEVA. "De CPU a GPU: acelerando los cálculos masivos", BEEVA, 4 de octubre de 2017. [En línea] URL disponible en: <https://www.beeva.com/beeva-view/tecnologia/de-cpu-a-gpu-acelerando-los-calculos-masivos/>

9.3 Escalabilidad

Hemos visto que uno de los factores clave es el efecto red y la capacidad de poder responder en un tiempo prácticamente cero a las nuevas necesidades de incremento de cómputo. En estas plataformas, solo tenemos que incorporar nuevas máquinas. Hay que tener en cuenta que esta infraestructura se compone de un número de máquinas de prestaciones medias; sumando muchas máquinas de prestaciones medias, conseguimos las mismas prestaciones que con una máquina muy grande, y con tolerancia a fallos. Esto nos permite escalar tanto en memoria como en almacenamiento de una manera muy sencilla sin tener que actualizar el código de nuestras aplicaciones.

Puede resultar sorprendente saber que la mayoría de los clústers de Big Data en realidad no son tan grandes. Mientras que mucha gente sabe que Yahoo está ejecutando más de 42.000 nodos en su entorno de Big Data, en 2013 el número promedio de nodos de un clúster de Big Data era poco más de 100, de acuerdo con Hadoop Wizard. El desafío de la escalabilidad tiene menos que ver con lo grandes que son ahora los clústers y más con cómo escalar con gracia para futuras implementaciones⁷.

La opción más idónea es que las organizaciones puedan acceder a un modelo en el cual las inversiones iniciales necesarias (CAPEX) para la compra de servidores, cabinas de discos, licencias de *software*, etc., se conviertan en gastos operacionales mensuales (OPEX) y puedan disponer en cada momento de la última tecnología sin riesgo de obsolescencia y en modo pago por uso⁸.

El *cloud* es una opción de infraestructura por la que muchas empresas optan para maximizar sus capacidades de escalabilidad futura.

⁷Michael Bushong. "Seis consideraciones para redes de Big Data", Tech Target, octubre de 2014. [En línea] URL disponible en: <https://searchdatacenter.techtarget.com/es/cronica/Seis-consideraciones-para-redes-de-big-data>

⁸Julián García Barbosa. "Cloud computing, el mejor camino para acceder al Big Data", Blog ThinkBig, 10 de junio de 2013. [En línea] URL disponible en: <https://aunclidclastic.blogthinkbig.com/cloud-computing-mejor-camino-para-acceder-al-big-data>

9.4 Competitividad

El uso del aprendizaje automático y la analítica avanzada nos permite conocer en profundidad a nuestros clientes y personalizar nuestros servicios para cada uno de ellos, es decir, ser una compañía centrada en el cliente en lugar de ser una compañía basada en servicios o productos.

Antes de poder aplicar estas tecnologías se requería una cantidad inmensa de recursos humanos, lo que hacía impensable poder centrarnos en el cliente. Ahora, sin embargo, es una gran oportunidad que crea ventajas competitivas sostenibles y nos diferencia de la competencia no solo en cuanto a servicios, sino también en costes por la productividad que podemos obtener de las aplicaciones basadas en aprendizaje automático, como hemos visto a lo largo de la unidad.

X. Resumen



Nos encontramos en un momento de cambio que ninguna generación anterior ha vivido, ni siquiera la generación de la Revolución Industrial. Existen cinco factores clave:

- Ley de Moore.
- Reducción del coste de almacenamiento y memoria.
- El proceso de digitalización y el cambio de hábitos de los consumidores, que son cada vez más digitales.
- El efecto red, que permite crecer de manera exponencial, con las implicaciones de agilidad e inmediatez. Es decir, debemos estar preparados para adaptar nuestros negocios de manera rápida a las necesidades de los clientes y para soportar un posible crecimiento exponencial.
- Aprendizaje automático como base de productividad y adaptación al cliente (nos permite definir servicios digitales centrados en cada cliente, personalización).

Todos estos factores tienen un elemento común, los datos. Si bien los datos con los que cuenta una empresa son el reflejo de su actividad, condicionada por sus productos, políticas comerciales, recursos humanos y las operaciones de la compañía, un dato en bruto no aporta ninguna información. Debemos transformar esos datos en conocimiento e inteligencia que nos proporcionen aquella información clave que nos permita tomar decisiones.

La gran mayoría de las compañías tienen una gran cantidad de datos, pero no todos los datos tienen el mismo valor siempre que se refieren a un problema concreto. Tenemos, por tanto, la siguiente clasificación de datos:

- **Datos disponibles.** Todos los datos con los que cuenta una empresa.
- **Datos inteligentes.** La fracción de los datos disponibles que son de utilidad para solucionar un problema o proveer un servicio.
- **Datos enriquecidos.** Aquellos datos inteligentes que se enriquecen con otros datos adquiridos de fuentes externas.

A la hora de analizar y usar los datos, tenemos cuatro tipos diferentes de analítica, que son un indicativo de la madurez analítica de la compañía:

- **Analítica descriptiva.** Se analizan los datos aplicando diferentes técnicas estadísticas para entender qué es lo que ha ocurrido.
- **Analítica de diagnóstico.** Es el siguiente paso y aplicamos técnicas de causa-efecto para entender el motivo por el que ha ocurrido un hecho.
- **Analítica predictiva.** Aplicando técnicas de aprendizaje automático e inferencia estadística, se intenta responder a qué es lo que va a ocurrir. Da respuesta a una pregunta basándose en los datos.
- **Analítica prescriptiva.** Se basa fundamentalmente en una técnica de aprendizaje profundo que se conoce como aprendizaje por refuerzo e intenta decirnos qué debemos hacer para que algo ocurra.

Las empresas se enfrentan al reto de adaptar sus organizaciones al nuevo contexto económico y para ello deben superar cuatro grandes desafíos:

- **Nuevo liderazgo.** Debemos encontrar nuevos líderes que impulsen el cambio y tengan visión digital y, más importante, de aplicación de aprendizaje automático.
- **Tecnología.** Debemos actualizar nuestros sistemas para adaptarlos al nuevo contexto y necesidades.
- **Talento.** Las empresas deben incorporar nuevos profesionales capacitados para el uso y despliegue de las nuevas tecnologías tanto de aprendizaje automático como de infraestructura Big Data.
- **Cultura empresarial.** Las empresas deben adaptar su cultura para poder responder a la velocidad necesaria a las nuevas reglas que impone el contexto actual.

Las empresas deben implantar la nueva cultura analítica estableciendo nuevas estructuras que les permitan, en función de su madurez, ir creciendo para convertirse en compañías orientadas al cliente y basadas en la analítica. Tenemos cuatro tipos:

- **Concentrado en un área funcional.** El equipo analítico reside en una única área funcional.
- **Modelo centralizado.** Se crea un equipo en una única ubicación geográfica y organizativa (o se hace corporativa la analítica del área funcional).
- **Centro de excelencia o competencia.** En este modelo, existe un equipo centralizado que da soporte a las áreas funcionales, pero estas tienen analistas propios, a los que sirve el centro de excelencia proporcionándoles consejo, tecnología, modelos y mejores prácticas.
- **Descentralizado.** Cada área funcional o geográfica tiene su propio equipo y herramientas analíticas.

Ejercicios

Caso práctico

Se reproduce a continuación un artículo de la página web Inside Big Data, “The Exponential Growth of Data”, del 16 de febrero de 2017 (<https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/>), referente al crecimiento de datos y su impacto en las tecnologías Big Data.



Noticias: The Exponential Growth of Data

The Intelligent Use of Big Data on an Industrial Scale

Several decades ago, Saudi Oil Minister Sheikh Yamani gained recognition for his insight into global development: “The Stone Age did not end for lack of stone, and the Oil Age will end long before the world runs out of oil.” Today, we live in what many call the Information Age, and we are in absolutely no danger of running out of information, particularly in data form. There is a general perception that we are overwhelmed with data, making the ability to store, process, analyze, interpret, consume, and act upon that data a primary concern. For large-scale, multi-national organizations and those in heavily regulated industries— such as finance, healthcare, or those covering multiple industry verticals — the situation becomes even more complex and challenging. Escalating data concerns are rampant in the Internet of Things (IoT) Age, during which growth of data is exceeding the capacity of traditional computing. The question then becomes, how do we consume those data sources and transform them into actionable information?

The Exponential Growth of Data

There are many sources that predict exponential data growth toward 2020 and beyond. Yet they are all in broad agreement that the size of the digital universe will double every two years at least, a 50-fold growth from 2010 to 2020. Human- and machine-generated data is experiencing an overall 10x faster growth rate than traditional business data, and machine data is increasing even more rapidly at 50x the growth rate.



The acquisition and analysis of data and its subsequent transformation into actionable insight is a complex workflow which extends beyond data centers, to the edge, and into the cloud in a seamless hybrid environment. The utilization of edge devices, in situ-computation and analysis, centralized storage and analysis, and deep learning methodologies which accelerate data processing at scale requires a new technological approach. Historically, data processing and analytics systems had specialized features for business analytics and high-performance computing (HPC) workloads. Yet with the advent of big data and industry standard x86-based computing, we are seeing a convergence in big compute, big data, and IoT for analytics. IDC research categorizes this convergence as high-performance data analytics (HPDA).

The HPDA market is at the center of big data analytics. The key factor driving the adoption of data-intensive computing is the need to rapidly analyze exploding volumes of data at the point of creation and at scale. An important consequence of this explosion is the need for users to adopt advanced data analytics technologies. Enterprises now have access to cheaper and more powerful computing platforms, and modern analytics software like Hadoop and Spark enable realtime analytics for a wide range of use cases, including fraud and anomaly detection, business intelligence, affinity marketing, product design and development, process automation, and personalized medicine. In addition to these software frameworks, implementing storage capacities and capabilities which enhance data flow, in-place analytics, and storage efficiency such as object storage and high-performance distributed file systems, is critical for effective scaling.

According to IDC's survey on the most important digital transformation projects, respondents cited cloud transformation/transition (66%), IoT (32%), and big data/cognitive solutions (27%) as key initiatives for big data usage and development. The cloud provides scalability, and the IoT forms the foundation for investments in big data and cognitive computing. IDC predicts that by 2020 50% of all business analytics software will incorporate prescriptive analytics built on cognitive computing technology, and the amount of high-value data will double, making 60% of information delivered to decision makers actionable.

Data Growth Challenges

The exploding volume and speed of data growth has introduced several challenges:

- System management and growing cluster complexity
- Data center power, cooling, and floor space limitations
- Storage, data movement, and management complexity
- Lack of support for heterogeneous environment and accelerators
- Significant shortage of skills to integrate and manage the big data ecosystem

Infrastructure Drives Improvement

Organizations are evaluating and implementing infrastructure to drive the following improvements:

- Manage growth and operational cost of big data infrastructure
- Provide elasticity and flexible capacity
- Ensure performance for diverse workloads
- Rapidly deploy and scale infrastructure
- Simplify management with Big Data as a Service (BDaaS)

In this document, our focus is on “industrializing” big data infrastructure—bringing operational maturity to the Hadoop data ecosystem, making it easier and cost-effective to deploy at enterprise scale, and moving companies from the proof of concept (PoC) stage into production-ready deployments.

Preguntas

1. En este artículo, se habla de un crecimiento exponencial de los datos hasta 2020 y más allá. ¿Cuáles crees que son los factores que pueden llevar a este rápido incremento de los datos?
2. En este artículo, las tecnologías Big Data se incluyen dentro del ámbito de *high-performance data analytics* (HPDA). ¿Por qué crees que las tecnologías Big Data son la clave para poder manejar este volumen de crecimiento acelerado de datos?
3. En este documento se habla de tecnologías HPDA y tecnologías *cloud*. ¿Cuándo crees que una empresa debe optar por una compra de infraestructura propia (*on-premise*) o moverse hacia el *cloud*?

Solución

1

En este artículo, se habla de un crecimiento exponencial de los datos hasta 2020 y más allá. ¿Cuáles crees que son los factores que pueden llevar a este rápido incremento de los datos?

El creciente volumen de información se debe a varios factores. Por ejemplo:

- ➔ **Web.** Toda la información de comportamiento del usuario en medios digitales —cada clic, cada acción— se mide y se recoge y, con ella, es posible conocer el comportamiento digital de los usuarios y, por tanto, hacerlo más eficiente para maximizar el retorno obtenido de ellos.
- ➔ **Redes sociales.** Uno de los medios que hoy proporcionan mayor volumen de información son las redes sociales. Todos los días se generan ingentes cantidades de información que, con una correcta explotación, muestran de forma clara las tendencias en opinión y comportamiento colectivo de sus usuarios.
- ➔ **Internet of Things.** Hoy en día, la gran mayoría de los dispositivos que tenemos en nuestro hogar tienen una IP. Las ciudades inteligentes, los coches inteligentes... Todos ellos son fuentes de información que vuelcan datos en tiempo real a grandes centros de información, donde, con un procesamiento eficiente, pueden mostrar pautas de comportamiento para optimizar su eficiencia, reducir la tasa de fallos y, en definitiva, optimizar su operativa.
- ➔ **M2M.** En este caso, nos referimos a toda la información generada por los *logs* de sistemas e infraestructuras. Un procesamiento inteligente de esta información permite hacer más eficiente la operativa diaria de estos sistemas en tiempo real, calcular indicadores tempranos de fallos y maximizar la utilización de estos sistemas.
- ➔ **Información no estructurada.** Hoy en día se estima que más del 80 % de la información disponible es no estructurada, ya sea texto, grabaciones de audio, imágenes, vídeos, etc. Ser capaces de procesar esta información y extraer indicadores de valor abre un abanico de nuevas posibilidades y capacidades para las empresas.

2

En este artículo, las tecnologías Big Data se incluyen dentro del ámbito de *high-performance data analytics* (HPDA). ¿Por qué crees que las tecnologías Big Data son la clave para poder manejar este volumen de crecimiento acelerado de datos?

El principal elemento que hace que las tecnologías Big Data sean una solución para la necesidad de procesamiento masivo de datos es la relación potencia-coste.

Big Data es una tecnología paralelizable, es decir, el escalado es horizontal. El escalado horizontal consiste en añadir infraestructura en paralelo. Con este esquema el incremento en potencia es lineal con respecto al coste. Por ejemplo, si tenemos un clúster de 10 máquinas, con un coste por máquina de 5.000 euros, y necesitamos duplicar la potencia, se añadirían 10 máquinas más al clúster, con un coste total de 50.000 € extra. Si necesitáramos volver a duplicar la potencia, haríamos otra vez lo mismo.

Otro tipo de tecnologías, con escalado horizontal, necesitan aumentar la potencia de los servidores. Y un servidor con potencia N y coste Y no sigue la norma de potencia $2*N$ es igual a $2*Y$ de coste, sino que el coste habitualmente es $K*Y$, donde K es un valor mayor de 2. Por tanto, la relación potencia-coste en un escalado vertical sigue una curva exponencial, no lineal como en el caso del Big Data.

Unido a esto, los sistemas de almacenamiento de datos en Big Data se basan en la premisa de usar muchos discos pero de bajo coste, replicando la información entre ellos. Esto hace que el coste de almacenamiento de información decrezca considerablemente.

Todo esto, unido, hace que el coste de almacenamiento y procesamiento haya bajado drásticamente, permitiendo, con el mismo presupuesto, procesar y trabajar sobre un volumen mucho mayor de información.

3

En este documento se habla de tecnologías HPDA y tecnologías *cloud*. ¿Cuándo crees que una empresa debe optar por una compra de infraestructura propia (*on-premise*) o moverse hacia el *cloud*?

Uno de los puntos principales de decisión cuando se trabaja con tecnologías HPDA es si realizar una compra de infraestructura u optar por una aproximación *cloud*. En función del tipo de proyecto y necesidades de procesamiento, es mejor optar por una u otra. A continuación, se detallan algunas de sus características:

	<i>On-premise</i>	<i>Cloud</i>
Pros	<ul style="list-style-type: none"> Habitualmente más barato para cargas no elásticas (de momento) Mejor rendimiento sobre el mismo <i>hardware</i> Control completo del sistema Separación física 	<ul style="list-style-type: none"> Modelo de costes <i>leasing</i>, más económico para cargas dinámicas Servicios gestionados Aplicaciones elásticas, adaptadas a la demanda en tiempo real Pago por uso
Contras	<ul style="list-style-type: none"> Servicios de infraestructura necesarios (instalación y mantenimiento) Alta barrera de entrada, conocimientos de infraestructura y Big Data Amortización 	<ul style="list-style-type: none"> Menor control del <i>hardware</i> Latencias en integraciones con sistemas <i>on-premise</i> Privacidad en <i>software</i>, no física <i>Lock-in</i> si no se usa un IaaS con distribuciones tipo Horton/Cloudera

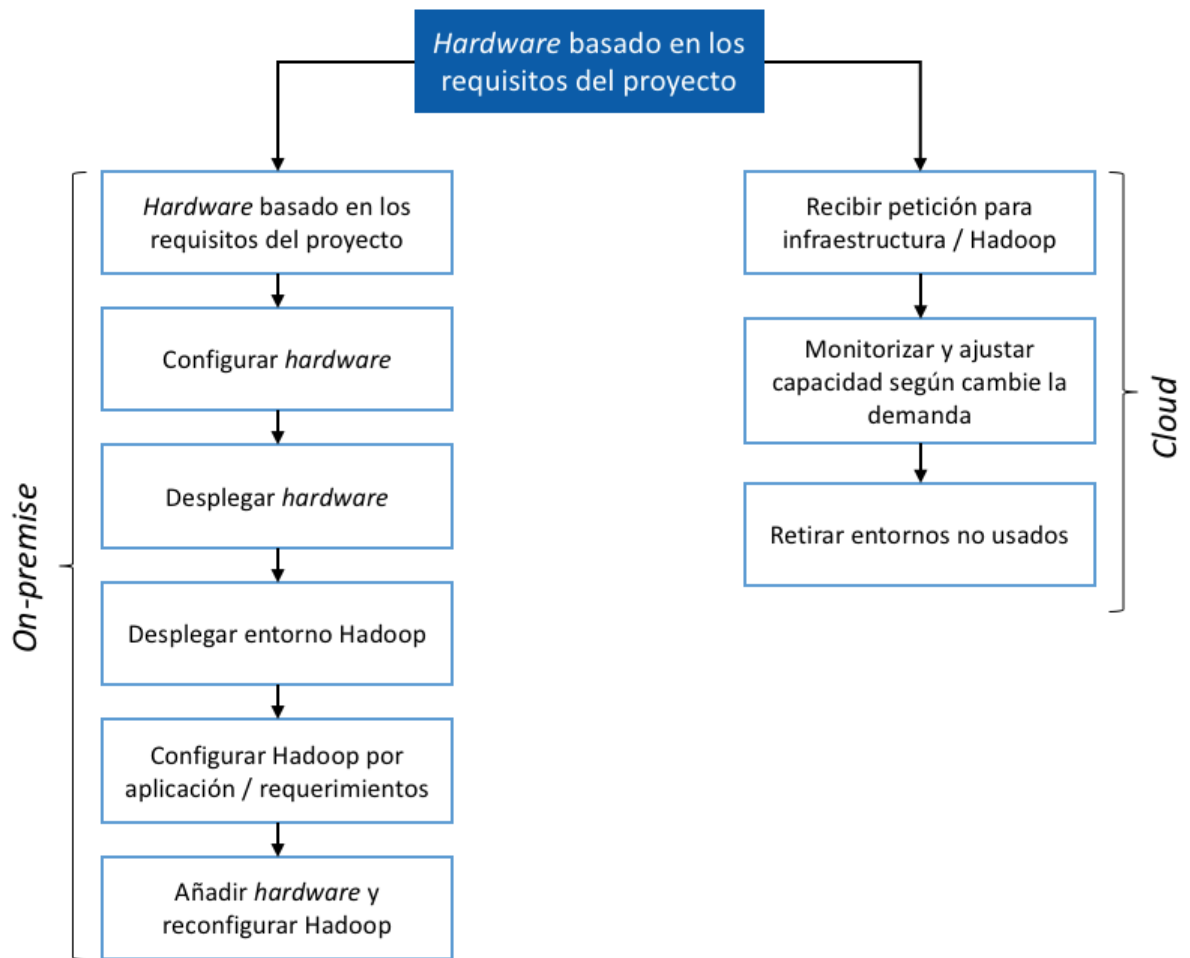


Figura 16. Infraestructura cloud vs. on-premise. Fuente: elaboración propia.

Recursos

Bibliografía

- ***Big Data and the Internet of Things: Enterprise Information Architecture for a New Age***: Stackowiak, R.; Licht, A.; Mantha, V.; Nagode, L. Apress; 2015.
- ***Big Data, Big Innovation: Enabling Competitive Differentiation through Business Analytics***: Stubbs, E. John Wiley & Sons; 2014.
- ***Big Data: A Revolution That Will Transform How We Live, Work, and Think***: Cukier, K.; Mayer-Schönberger, V. John Murray Publishers Ltd.: 2013.
- ***Big Data: Understanding How Data Powers Big Business***: Schmarzo, B. John Wiley & Sons: 2013.
- ***Gestión de la información en las organizaciones. Principios, conceptos y aplicaciones***: Ponjuán Dante, G. Santiago de Chile: CECAPI; 1998.
- ***Analytics Across the Enterprise: How IBM Realizes Business Value from Big Data and Analytics***: Dietrich, B.; Plachy, E.; Norton, M. IBM Press, 2014; May 15.
- ***Big Data: The Management Revolution***: McAfee, A.; Brynjolfsson, E. *Harvard Business Review*, 2012.

Glosario.

- **BI**: Business Intelligence, son sistemas tradicionales de procesamiento de información y explotación mediante informes y cuadros de mando.
- **Big Data**: Tecnología para el procesado masivo y ágil de datos, que se ha convertido en el acelerador tecnológico para la adopción por parte de los negocios de proyectos en Analytics, minimizando costes y maximizando el ROI.
- **Business Analytics**: Rama de las tecnologías de la información centrada en la aportación de valor a las áreas de negocio por medio del análisis de datos mediante técnicas estadísticas, de machine learning y de inteligencia artificial.
- **CAPEX**: Inversiones dentro de una empresa que crean beneficios.
- **CEO**: Chief Executive Officer o director ejecutivo.
- **Cloud**: Infraestructura virtual, flexible y con opciones de pago por uso.
- **CPU**: Unidad central de procesamiento, elemento principal en equipos informáticos encargado del procesamiento de información.
- **Cross-selling**: Proceso de venta enfocado a vender nuevos productos o servicios al cliente.
- **GPU**: Unidad gráfica de procesamiento, coprocesador dedicado al procesamiento de gráficos u operaciones en coma flotante para aligerar la carga de trabajo de la CPU.
- **IoT**: Se refiere a la tendencia a conectar mediante internet elementos cotidianos del día a día. Se trata por tanto de cualquier elemento con una IP y capacidad para enviar y transmitir datos.
- **M2M**: Machine-2-machine, se refiere a un concepto genérico que se refiere al intercambio de información entre sistemas tecnológicos sin intervención humana.

- ➔ **On-premise:** Sistemas hardware físicos, instalados y operados dentro de las instalaciones de la organización.
- ➔ **OPEX:** Coste permanente para el funcionamiento de la empresa.
- ➔ **Revenue:** Ingresos obtenidos.
- ➔ **ROI:** Return of investment. Es un índice financiero que mide y compara el beneficio o la utilidad obtenida con relación a la inversión realizada.
- ➔ **Smart city:** Es el término acuñado para las ciudades inteligentes, que aportan nuevas capacidades y servicios de valor mediante la sensorización y monitorización de parámetros de la ciudad.
- ➔ **SSD:** Unidad de estado sólido, es un tipo de dispositivo de almacenamiento de datos que, en lugar de basarse en discos magnéticos, hace uso de memoria no volátil como la flash. Son mucho más rápidos en ciertas condiciones que los discos magnéticos.
- ➔ **Time-to-market:** Tiempo desde la concepción de una idea hasta la salida de un producto o servicio al mercado.
- ➔ **Up-selling:** Proceso de venta enfocado a generar mayores ingresos sobre un tipo de servicio o producto ya adquirido por el cliente.