



Schmarzo, Bill. *Big Data: El poder de los datos*. Madrid: Anaya Multimedia, 2014, 245 págs. ISBN: 978-84415-3576-3.

El autor de esta monografía trata de ayudarnos a conocer ¿Qué es lo que nos frena con respecto a los Big Data? y ¿Qué hacer?

Según Bill Schmarzo los Big Data sirven para hacer dinero y su misión es mostrar cómo utilizar estos conocimientos prácticos sobre clientes, operaciones y productos para hacerlo. Desde una perspectiva pragmática y práctica con técnicas, herramientas y metodologías.

Para B.Schmarzo: “*Sólo estamos a un paso en lo que a tecnología se refiere de dar el gran salto en nuestra experiencia con los Big Data*”.

El índice de esta obra se compone de trece capítulos. *En el primer capítulo, los autores presentan la oportunidad de negocio que suponen los Big Data*. La revolución de los Big Data proporciona datos estructurados, semiestructurados (archivos generados por sensores), no estructurados (es decir, documentos de texto, publicaciones en medios sociales, comentarios, etc.) Lo que ha hecho necesaria una innovación tecnológica propiciada por empresas de medios digitales como: Google, Yahoo, Facebook. Las empresas que adoptan los Big Data como medio de transformación pasan a un enfoque predictivo y previsor. Para ayudarnos a entender este cambio nos presenta el caso Walmart que modificó la relación proveedor-minorista, al analizar los datos de los puntos de venta. Esta transformación se vincula al Índice de Madurez de los Modelos de Negocio de los Big Data. Que nos ayuda a plantear las siguientes cuestiones:

- 1) *¿Hasta dónde nos pueden llevar los Big Data, desde el punto de vista empresarial?*
- 2) *¿Cuál podría ser el objetivo final?*
- 3) *¿Cómo me puedo comparar con los demás con respecto a la adopción de los Big Data? ¿Cómo estímulo para los negocios?*
- 4) *¿Hasta dónde puedo llevar los Big Data para potenciar (o incluso transformar mis procesos de creación de valor?*

Este Índice de Madurez de los Modelos de Negocio de los Big Data cuenta con 5 fases: 1) Monitorización Empresarial. Llamada también “*gestión del rendimiento empresarial*”. Identifica el área de rendimiento inferior o superior. Captura los KPI (*Key Performance Indicators*), indicadores clave del desempeño), las dimensiones, las métricas, etc. 2) La Fase de Perspectiva Empresarial. Realiza la monitorización

empresarial, al aprovechar las nuevas fuentes de datos sin estructurar mediante estadísticas avanzadas, analítica en tiempo real, identifica los conocimientos materiales significativos y aplicables que se pueden integrar en los procesos claves de la empresa. 3) La Fase de Optimización Empresarial. Nivel en el que las organizaciones utilizan análisis integrados para optimizar partes de sus operaciones empresariales. 4) La Fase de Monetización de datos. Las organizaciones buscan sacarle partido a los Big Data para tener la posibilidad de obtener beneficios. Crear productos inteligentes. O sacar partido a conocimientos que se puedan convertir en acciones personalizadas y recomendaciones. Las tendencias del cliente para mejorar la relación con éste y sopesar cuál es su experiencia. 5) La Fase de Metamorfosis Empresarial es el objetivo final. Aprovechan los conocimientos que están recopilando para transformar sus modelos de negocio en nuevos servicios y nuevos mercados. Pero, ¿cómo se realiza? Esto nos lleva según los autores a 3 pasos:

- 1) Dedicar tiempo a investigar y hacer un seguimiento de nuestros clientes/usuarios para entender que soluciones buscan. Centrarse en los que intentan conseguir.
- 2) Conocer a los agentes potenciales del ecosistema (los desarrolladores) y cómo podrían ganar dinero con nuestra plataforma. Reunirse con ellos. Sacar ideas y priorizar sus posibilidades de convertir los datos en dinero.
- 3) Como proveedores de la plataforma, centrarnos en el desarrollo del producto, el marketing y la búsqueda de asociaciones para cerciorarnos de que la plataforma es: Fácil de cara al marketing, las ventas y el soporte. Es escalable y fiable en relación a su disponibilidad, ampliación, almacenamiento de datos y capacidad de procesamiento. Posee todas las herramientas, el procesamiento de datos, las aptitudes analíticas y las capacidades móviles. Simplifica el modo en que terceras partes hacen dinero (contratos, términos y condiciones, etc.) Permite recopilar y analizar los datos del uso y funcionamiento del producto para mejorar la interacción. Ayuda a los desarrolladores a optimizar sus operaciones empresariales.

De acuerdo con la evolución del Índice de Madurez se podrán apreciar tres cambios culturales: 1) *Los datos pasan a ser un recurso.* 2) *Las analíticas, los algoritmos y modelos se convierten en un modelo de propiedad intelectual que hay que gestionar, alimentar e incluso proteger.* 3) *La organización se siente cada vez más cómoda al tomar decisiones basadas en los datos y las analíticas.*

En el *capítulo segundo*, nos exponen una breve historia de los Big Data. Aludiendo a los escáneres TPV que dieron lugar a esta primera revolución de los Big Data. Cuyo volumen fue creciendo hasta llegar a terabytes de datos, reemplazando a los datos procedentes de auditorías bimestrales que se empleaban para la toma de decisiones. El volumen, la diversidad, y la velocidad de los TPV dejaron obsoletas la gestión de los datos y tecnologías analíticas existentes. El software de igual manera tuvo que evolucionar a tecnologías como las plataformas especializadas en gestión de datos (Red Brick, Teradata, etc.)

Las empresas capaces de crear nuevas aplicaciones de negocio basadas en la analítica, como la administración de categorías y la predicción basada en la

demanda han sido las ganadoras. Al igual que los minoristas con acceso inmediato a los escáneres de TPV y con datos de una clientela fiel han analizado comportamientos, lo que les ha permitido cambiar el equilibrio entre la industria y los fabricantes al prescribir las condiciones sobre los productos (precios, etc.)

En el *capítulo tercero*, los autores nos exponen el Impacto Empresarial de los Big Data. Para Schmarzo los Big Data permiten a las empresas responder a preguntas que antes no podían, tipo: 1) *¿Quién es mi cliente más valioso?* 2) *¿Cuáles son mis productos más importantes?* 3) *¿Cuáles de mis campañas tienen más éxito?* 4) *¿Cuáles de mis canales funcionan mejor?* 5) *¿Cuáles son mis empleados más eficaces?*

Cuanto más se piense es más evidente que responderlas no es sencillo. Gracias al conocimiento que brindan los Big Data, se pueden llevar este tipo de preguntas al siguiente nivel de comprensión que nos permitirá dar lugar a respuestas más precisas y proporcionar un entendimiento más completo de nuestra clientela/usuarios, nuestros productos y las operaciones que pueden tener repercusión en varios aspectos del negocio, como: 1) *Las promociones*, campañas más eficaces, o atraer tráfico a la web. 2) *El marketing*. 3) *Las ventas* para optimizar la ubicación de recursos de ventas escasos frente a las mejores opciones de venta, etc. 4) *Las compras*, para saber que proveedores tienen los costes de distribución más ajustados para los productos de alta calidad. 5) *La producción*. 6) *Recursos humanos*, para identificar las características y el comportamiento de los empleados más eficaces y con mayor éxito.

Este impacto empresarial se traduce en la monetización de datos. Para ello nos proponen un ejemplo basado en los medios digitales: Yahoo, Google, Facebook, Twitter que trabajan para ello. *Pero, ¿cuáles son los procesos que deben seguir estas empresas de medios digitales para saber cómo monetizar sus recursos de datos?* Nos proponen conocer dos conceptos clave: 1) *¿Quiénes son mis clientes objetivo (personas influyentes) y por qué soluciones empresariales estarían dispuestos a pagar?* 2) *¿De qué activos de datos dispongo (o podría disponer)?*

Para iniciar el proceso de monetización sugieren poseer un sólido conocimiento de estos. Las empresas de medios digitales están dirigidas a estos tres tipos de clientes: 1) planificadores y compradores de medios, 2) directores de campaña y 3) ejecutivos de los medios digitales. Estos compran “soluciones” como las siguientes: 1) Públicas; 2) Inventarios; 3) Resultados de medidas, como el coste por miles (CPM), etc.

Estas empresas evalúan la amplitud, profundidad y calidad de sus activos de datos como: 1) Los visitantes y sus *datos demográficos o psico-demográficos* asociados, y lo que sabemos de su comportamiento. 2) *Las propiedades y el tipo de contenido que proporcionan los diversos tipos de anuncios* que poseen en páginas como Yahoo, etc. 3) *Actividades que realizan los visitantes sobre estas propiedades* (si han visualizado en pantalla, han pasado el ratón sobre un anuncio, han hecho clic sobre éste o introducido una palabra clave) incluyendo la frecuencia, cuánto hace y en qué secuencia.

Este tipo de proceso también debería incluir qué datos adicionales se podrían recopilar mediante la adquisición de datos. Una vez recopilada toda esta

información es necesario tener capacidad de procesamiento y los conocimientos analíticos avanzados para perfilar, segmentar y empaquetar a estos visitantes en un formato de público que los anunciantes y los agentes quieran comprar. Los pasos que se deben seguir son: 1) *Identificar a los clientes objetivos y las soluciones que sean.* 2) *Inventariar y evaluar nuestros conocimientos necesarios.* 3) *Conocer los procesos de incorporación, transformación, limpiado, clasificación, enriquecimiento y análisis de datos necesarios para transformar nuestros activos en soluciones empresariales.*

El *impacto de los Big Data* en la organización (cap.4) se resume en el cambio organizacional o la transformación necesaria para explotar y soportar las posibilidades que ofrecen. Conlleva pasar de la Inteligencia Empresarial a la Ciencia de los Datos. La Ciencia de los Datos se suele asociar con el pensamiento en el futuro. Sin embargo, algunas de las implementaciones de la Inteligencia Empresarial se usan para ayudar a tomar decisiones de futuro.

Una de las principales diferencias entre las analíticas de la Inteligencia Empresarial y los científicos de datos es el entorno de trabajo. Los primeros trabajan con datos Bases de Datos muy estructuradas, en entornos orientados a la producción. Los científicos de datos crean un sandbox analítico independiente en el que caben cualquier cosa (fuentes de datos tanto internas como externas) y con el nivel de granularidad o el historial que necesiten. El científico de datos tiene libertad total para transformar, desarrollar, probar y refinar modelos analíticos.

Las mejores organizaciones definen un proceso, los roles, responsabilidades y expectativas. Para ello se valen del ciclo vital de la analítica de los datos que facilita la colaboración con: 1) *El usuario profesional.* Es el único que conoce las preguntas que intentan responder y las decisiones que va a intentar tomar. 2) *El administrador de la Base de Datos* es el responsable de datos. 3) *El científico de datos* es responsable de explotar los datos de la organización para descubrir nuevos conocimientos. 4) *El proceso analítico* completa el ciclo volviendo a las empresas usuarias, informes, gráficas y los conocimientos analíticos resultantes para dirigir sus negocios.

No obstante, la naturaleza de los roles variará dependiendo de la organización o el proyecto. En relación a estos nos proponen crear: El equipo de la interacción de usuarios. El rol de administrador senior y liberar la creatividad empresarial.

Respecto a cómo funciona *la teoría de decisiones nos lo explica en el capítulo 5.* En este proceso hay que salvar *algunos escollos como el factor humano* en la toma de decisiones y en concreto conocer *algunas de las trampas como:*

- 1) *El exceso de confianza* al sobrevalorar su experiencia en un mercado existente cuando se pasan a un nuevo o lanzar un producto diferente. Para evitarlo recomiendan: Implementar procesos estructurados de toma de decisiones; Investigar para reunir todos los hechos y entender los riesgos; Utilizar comisiones revisoras.
- 2) *El sesgo de la obstinación.* Aferrarse a un hecho y usarlo como punto de referencia en las decisiones. Para evitarla recomiendan: Consultar la opinión de varios expertos; Desarrollar modelos para entender las dinámicas y

relaciones institucionales, empresariales; Planificar en base a un proceso continuo, no a un evento.

- 3) *La aversión al riesgo*. El rechazo a aceptar una oferta con una recompensa incierta en vez de una con una recompensa más segura, aunque más pobre. Para evitarlo: Emplear un proceso estructurado que recopile y pondere tanto los riesgos como el coste de oportunidad de no hacer nada. Utilizar expertos para minimizar los sesgos en el criterio.
- 4) *No saber cuáles son los gastos no recuperables*. Para evitarlo recomiendan: Crear modelos empresariales que consideren los gastos no recuperables.; Asegurarse de que el análisis sólo tenga en cuenta los nuevos costes incrementales; Utilizar expertos para minimizar el criterio.
- 5) *La forma de presentar los datos*. El modo en que se enuncia o se presenta una opción puede repercutir en la decisión. Para evitarla recomiendan: Cerciorarse de que los modelos de decisión toman la misma referencia y se basan en las mismas suposiciones.; Crear modelos desde muchas perspectivas para asegurarse de que a las variables se les asigna el peso adecuado.

Con la analítica que permiten los Big Data, la interfaz de usuario, se podrá centrar en dos tipos de información: los conocimientos. Los *conocimientos* pueden aprovecharse para identificar situaciones que se salen de los límites normales. Las *recomendaciones* son acciones específicas que se generan en base a un análisis detallado de los conocimientos y el estado actual de la empresa. Nos sugieren que la interfaz podría empezar por estos Conocimientos y Recomendaciones, ordenados en función del impacto potencial.

¿Cómo se crea una estrategia para los Big Data? En el capítulo 6 nos ofrecen un proceso y un marco de trabajo plasmado en un documento que ha de ser: Conciso 1 página. Claro define lo que tiene que hacer la organización, las personas. Relevante para los grupos de interés, al partir y centrarse en el proceso de apoyo a la estrategia empresarial global de la organización. El documento ha de contener las siguientes secciones:

1. Estrategia empresarial (título del documento).
2. Iniciativas empresariales en las que se sustenta.
3. Resultados y factores críticos de éxito (FCE) (definen lo que hay que hacer). Los resultados definen el estado final deseado o ideal.
4. Tareas: Las tareas específicas, las tareas clave. Es el ¿cómo se hace? Es el nivel en el que se asignan, miden las tareas personales y los objetivos de la gestión.
5. Fuentes de datos: Fuentes de datos clave para sustentar la estrategia empresarial y las iniciativas empresariales clave de apoyo.

Una vez que se ha realizado es el momento de pasar a la acción. El documento divide los requisitos tecnológicos en los siguientes:

- 1) Las métricas, indicadores clave de rendimiento con los que se medirá el progreso y el éxito.
- 2) Las preguntas, las métricas y el flujo de decisiones necesario.
- 3) Los algoritmos analíticos y los requisitos de modelado y transformación de datos en que se apoyarán.
- 4) Las fuentes de datos de apoyo, medidas, dimensionalidad, atributos dimensionales, granularidad, frecuencia de actualización, almacenamiento y métodos de acceso.

Tras este proceso se podrá definir la arquitectura analítica y la tecnología, incluyendo la administración de la base de datos, el enriquecimiento de datos, etc.

Para saber cómo aprovechar estos conocimientos para transformar a nuestros clientes, usuarios y nuestra interacción con ellos. Además de aprender a mirar fuera de nuestra organización para determinar el uso a hacer de los Big Data (cap.8) nos proponen este proceso que consta:

- Entender los objetivos de los clientes para que haya comunicación entre ellos y nuestra organización
- Captar las decisiones que los clientes necesitan tomar con respecto a sus objetivos para mejorar la efectividad de sus interacciones con nuestra organización, productos y servicio.
- Identificar que información necesitan los usuarios para basar las decisiones que necesitan tomar.

Del mismo modo, también nos exponen como descubrir y aprovechar los conocimientos sobre los clientes/usuarios encaminados a reestructurar los procesos de creación de valor. A través de huella digital por toda la red. Esta huella proporciona unos conocimientos valiosísimos dentro de sus áreas de interés (“me gusta”), pasiones (defensores y partidarios), asociaciones (grupos a los que pertenecen oficialmente) y afiliaciones (causas en las que creen) y que se pueden utilizar para mejorar aspectos de la comunicación con el cliente.

Estos conocimientos nos indica Schmarzo pueden tener repercusión en cualquier aspecto de la comunicación con el cliente: cómo creamos perfiles y segmentamos, como gestionamos a los defensores y partidarios. El desafío es cómo puede aprovechar su organización los conocimientos para optimizar todos los puntos de interacción con el cliente que comprenden el ciclo vital.

Las organizaciones necesitan complementar datos para descubrir conocimientos aplicables u obtener métricas financieras fundamentales, como el valor del cliente (LTV) y la probabilidad de recomendación (LTR).

Pero *¿Cómo podemos saber identificar los casos de uso de los Big Data?* La respuesta nos la presentan en el *cap.9* en el que se introduce una metodología demostrada y aplicada: El proceso de visualización que se resume en estos pasos:

1. Investigaciones y entrevistas para entender las iniciativas empresariales objetivo o el proceso empresarial.

2. Preparación de los datos y desarrollo de analíticas específicas para el cliente/usuario.
3. Ejercicios de visualización para transmitir el “universo de posibilidades”
4. Lluvia o tormenta de ideas y priorización de los casos de uso de los Big Data.
5. Recopilar los riesgos de implementación y los factores de creación de valor empresarial.

Posteriormente nos introduce el proceso basado en la matriz de priorización. Lo que se busca es el valor empresarial relativo de cada caso de uso y alguna explicación de los participantes acerca de la colocación del caso de uso. Nos expone el proceso, así como las trampas y por ello recomiendan evitar los casos de uso que se adentran en la “*zona de expectativas mal gestionadas*”, la “*zona de decepciones*”, o la “*zona de acción limitada*”. Para finalizar con una exposición sobre el uso de los modelos de interacción con el usuario para estimular el proceso de visualización.

El problema de los Big Data como afirman los autores es que no existe una solución técnicamente brillante. No basta con instalar una herramienta como Hadoop y dar por hecho que va a proporcionar una solución de Big Data. Es necesario el uso de lo que denominan Ingeniería de Soluciones (cap.10) y que se define: “*Un proceso para identificar y descomponer las principales iniciativas empresariales en las capacidades empresariales y tecnologías de apoyo que las componen, para poder dar una base a la toma de decisiones y los intentos de monetización en la organización*”. Este proceso abarca lo siguiente pasos:

1. Entender cómo hace dinero la organización.
2. Identificar las principales iniciativas empresariales de la organización.
3. Descomponer la iniciativa empresarial en casos de uso.
4. Evaluar los casos de uso.
5. Diseñar e implementar la solución de Big Data.

Para Schmarzo la ingeniería de soluciones se irá volviendo más importante conforme evolucione la cantidad y variedad de los datos, las capacidades tecnológicas y las aplicaciones para móviles, lo que nos explica a través de diversos ejemplos como el de analíticas de comportamiento del cliente, etc. Para terminar aclarándonos como se ha de leer el informe anual para identificar oportunidades de negocio de los Big Data y que nos ejemplifica para una empresa de servicios, entre otros.

La introducción a las tecnologías de los Big Data la realiza en el cap. 11. En el que se presentan brevemente algunas de las principales herramientas basadas en una nueva arquitectura de datos y analítica de escalabilidad horizontal basada en código abierto. Surge Apache Hadoop, Hadoop MapReduce, Apache Hive, Apache Hbase, Pig. Así como nos explica cuales son las nuevas herramientas analíticas como R, Lenguaje de programación y estadísticas de código abierto. O los nuevos algoritmos analíticos, entre ellos la matriz de confusión, etc, y continuar con la

explicación de cómo introducir los big data en el mundo del almacenamiento de datos tradicional a través del enriquecimiento de datos, la federación de datos, o el modelado de datos. Las Arquitecturas MPP, las Analíticas in database, y la Computación en la nube.

El viaje hacia los Big Data (cap.12) nos lo resumen en un storymap traducido en el presentan las principales observaciones y estrategias. Viaje marcado por el crecimiento de los datos que implica oportunidades de negocio para que los métodos y tecnologías tradicionales no sean suficientes. Por ello nos vuelve a hablar del Índice de Madurez del Modelo de Negocio de los Big Data. Y para que haya éxito en el viaje es necesario la colaboración entre la empresa y la TI junto con una aplicación de los conocimientos obre Big Data que a su vez impulsan el proceso de creación de valor.

Los autores finalizan la monografía con un *llamamiento a la acción (cap.13)* en el que recomiendan: identificar las principales iniciativas empresariales de la organización. Colaboración entre empresas y TI. Formalizar el proceso e visualización. Emplear modelos para impulsar el proceso creativo. Conocer su tecnología y las posibilidades de su arquitectura. Ampliar los conocimientos empresariales internos. Descubrir nuevas oportunidades de monetización. Así como las ramificaciones de la organización.

Lectura recomendable que nos introduce en este nuevo concepto cuya aplicación ya se está analizado por autores de la disciplina.

Ana Nieves Millán Reyes
Universidad de Granada
anmillan@ugr.es