

Y de repente Big Data

Fernando Arencibia¹

Resumen

La cada día mayor facilidad de acceso a las fuentes de información, la variedad de entornos en las que esta se genera y se procesa y el uso que de la misma se haga y el valor añadido que se pueda aportar a las empresas son elementos que cada vez tienen una mayor importancia en nuestro entorno de negocios no solo como fórmulas de mejora de rentabilidad sino como medio para la creación y mantenimiento de barreras competitivas que nos diferencien de los competidores. El acceso a los datos, su conversión en información, el conocimiento que de este se deriva y la traslación a la acción no pueden dejarse ya únicamente en operadores humanos sino que, haciendo uso de las capacidades que ofrece la tecnología, se puede acceder a un entorno competitivo mucho más flexible para dar respuesta a modelos de máxima exigencia. Este cuaderno nos ilustra de manera sencilla las principales variables que deberían ser tenidas en consideración en nuestras empresas.

Desde los tiempos de las primeras civilizaciones y probablemente incluso antes, el hombre ha intentado recolectar datos sobre los más variados temas de su entorno y su propia actividad, para, después de contrastarlos y analizarlos, poder utilizarlos para enfrentarse en las mejores condiciones posibles a los desafíos cotidianos: esto es, usamos de forma casi natural la información de que disponemos como generador de ventajas competitivas, bien frente a otros miembros de nuestra especie o bien frente al entorno hostil donde habitamos. Por los resultados obtenidos como especie, parece que la cosa, por ahora, al menos comparándonos con otros animales ya extintos, nos ha ido bastante bien.

Volviendo a nuestros días, si nos aproximamos al mundo empresarial, las compañías acumulan grandes cantidades de datos de su quehacer cotidiano, y conceptos técnicos como EDW (Data Warehouse corporativo) o Inteligencia de Negocio son de uso común. Esto significa que seguimos en el camino de intentar utilizar la información que tenemos transformándola en

¹ Fundador y Director General de HabberTec Business & Intelligence. Contacto: fernando.arencibia@habber.com

ventajas para nuestra actividad, en este caso en los negocios, y lo venimos haciendo de forma explícita y asignando presupuestos específicos desde hace más de veinticinco años con ayuda de las tecnologías de la información que han suministrado abundantes medios específicos para lograrlo.

Nada aparentemente nuevo hasta aquí ¿Por qué entonces tanto revuelo con este “revolucionario” concepto de BIG DATA? ¿Será cierto que supone un cambio de paradigma en nuestra relación con la información? Veamos si podemos encontrar un poco de luz.

¿Algo nuevo bajo el sol?

En estos momentos confluyen las siguientes circunstancias que desde mi punto de vista son claves para que estemos en este estado de excitación con BIG DATA:

I. Avalancha de información:

De forma ya ampliamente consensuada, por otra parte casi lo único consensuado que hay en este tema, se habla de que BIG DATA va asociado a tres “V” que lo definen, a saber: Volumen, Variedad y Velocidad. Todo respecto a la información, se entiende. Repasémoslas.

a. Volumen:

- En estos momentos, si miramos a la Red de Actividades de cualquier empresa, la práctica totalidad de las tareas se realizan apoyadas de una u otra forma en un sistema informático: correo, hojas de cálculo, sistemas ERP, sistemas CRM, sistemas de gestión logística, Call Centers, páginas web, sistemas transaccionales de soporte de la operación,... etc. Esto cada día que pasa acumula información sobre lo que hacemos, y cada vez es más cantidad al aumentar el número de estos sistemas y su complejidad. Los Datawarehouse clásicos utilizados para guardar toda esta información y después explotarla con sistemas analíticos, pueden hoy día superar fácilmente en compañías grandes varias decenas e incluso centenas de Terabytes sin ningún inconveniente, cuando hace menos de ocho años para la mayoría el Terabyte o como máximo la decena marcaba un límite impensable.
- También podemos comprar información a proveedores externos, como por ejemplo información sociodemográfica sobre nuestros clientes. Cada vez hay más información disponible para enriquecer la que ya tenemos.
- Desde hace algunos años, también se ha incorporado con fuerza la idea de controlar la Red de Actividades mediante la identificación de

los procesos de la empresa y su correcta gestión, de modo que sepamos no sólo lo que hacemos sino como lo hacemos, de ahí que se estén introduciendo cada vez más sistemas BPM (Business Process Management) que guarda ingente cantidad de información explicativa del cómo hacemos las cosas (cuanto tardamos, que equipos son más productivos, donde están los cuellos de botella,...).

- En esta misma línea, nos encontramos con la introducción de los sistemas BAM (Business Activity Monitoring) que permiten mediante la colocación de sensores de distintos tipos y el tratamiento de los datos en tiempo real recoger grandes cantidades de información que nos ayude a monitorizar las interioridades de nuestros negocios.
- Otra fuente ingente de información es la proveniente de los comentarios que podamos encontrar en nuestra web, o en los canales propios de Redes Sociales (Facebook, Twitter,...) o lo que registran los Contact Centers en su atención a clientes.
- Para completar la tarta nos encontramos con la enorme cantidad de información existente sobre nuestro negocio (o negocios competidores, o sustitutivos, u otras cuestiones relacionadas e importantes para nosotros) en lo que podemos llamar Social Web (Social Media & Web): webs, RSS, Blogs, Facebook, Twitter, Youtube, Twenty,....

No os quiero aburrir con números, pero considerad que en internet cada minuto se mueven unos 650 Gbytes de información, que diariamente se mandan 150 millones de tweets, o que las RFDI, que sustituirán a los códigos de barra pueden aportar unas 900 veces más información que estos últimos, o que se suben 35 horas de video cada minuto a You tube. Si de verdad estáis interesados en cifras sólo tenéis que ir al propio You tube y visionar cualquiera de los videos que hay sobre el crecimiento de la información en internet. Realmente asombroso.

b. Variedad:

- Cuando en el punto anterior hemos descrito los múltiples orígenes de la información que tenemos disponible actualmente, mirando a su procedencia y su formato podemos clasificarla desde dos puntos de vista:
 - i. Interna o Externa: Según la produzcamos en la empresa o venga del exterior.
 - ii. Estructurada o Desestructurada: Según sea información cuya estructura permita tratarla directamente con los sistemas de información (generalmente en formato de base de datos relacionales, almacenada en atributos, registros y tablas) o si

es información en lenguaje natural como puede ser una queja enviada al Contact center o un mensaje de twitter. Esta última es uno de los grande desafíos que se presentan en estos momentos, porque para sacarle partido debemos ser capaces no solo de capturarla y almacenarla, sino de llegar a una comprensión semántica de cada comentario, y ser capaces de responder a cuestiones tales como: ¿De qué está hablando este texto? ¿Es positivo o negativo? ¿A qué área de la empresa hace referencia?... En general al responder a estas preguntas lo que estamos intentando es pasar de información desestructurada a información estructurada, que es la que podemos utilizar para conseguir nuestros objetivos. Es como en una encuesta, en la que el encuestador recibe del encuestado mucha información desestructurada y con la ayuda de su cuestionario la transforma en información estructurada y utilizable.

c. Velocidad:

- Aquí también podemos encontrar dos puntos de vista claros: por una parte tenemos velocidad en la acumulación de la información, por lo continuo de su llegada que muchas veces nos obliga a tratar con flujos de datos en tiempo real (como el que nos envían los sensores), o la rapidez con la que podemos acumular mensajes en redes sociales sobre un tema muy caliente de opinión. Por otra parte tenemos la velocidad con que la información pierde su vigencia o “frescura”, por lo que si no somos capaces de que nuestros sistemas analíticos se adapten a esto podemos estar siempre mirando a preciosos indicadores que se han quedado obsoletos antes de poder ser útiles.

II. Evolución de las tecnologías de la Información:

- **Procesamiento y almacenamiento:** A nadie le es ajeno el impresionante desarrollo tanto de la capacidad de proceso como de la capacidad de almacenamiento que experimentan los ordenadores cada año: con mucho menor coste se consiguen muchas más prestaciones. En estos momentos la capacidad es tal que comenzamos a manejar problemas de análisis on line manipulando Terabytres de información sin que encontremos impedimento alguno en la potencia de los ordenadores, tanto grandes servidores como equipos de uso personal.
- **Software Específico:** En estos últimos años han aparecido el ecosistema Hadoop basado en software libre, que incluye conceptos como Hadoop

Distributed File System, Map Reduce, Hive, HBase, etc que pretenden dar solución al almacenamiento y análisis de información desestructurada. Esta tecnología ha sido ampliamente extendida y hoy en día muchos fabricantes de sistemas venden sus adaptaciones de estos productos. También las bases de datos relacionales analíticas usadas como base para Datawarehouse han aumentado muchísimo su número y sus capacidades, de hecho no hay ni un gran fabricante que no tenga su propio catálogo de producto para este segmento (Exadata de Oracle, Netezza de IBM, Teradata,...), además de productos como Hana de SAP, que con sistemas de trabajo “in memoy” presumen de prestaciones realmente sorprendentes.

- **Movilidad:** Nuestro mundo actual es móvil, cada vez en mayor porcentaje las comunicaciones y las transferencias de información se realizan desde o hacia un dispositivo móvil como teléfonos inteligentes o tabletas. Con esto prácticamente no hay tiempos muertos en la capacidad de las personas para ser proveedores o consumidores de información, lo que hace que el flujo de la misma sea espectacular, además ya no se limita a texto, también se incluyen imágenes o videos con el consiguiente incremento de volumen.

III. Modelos de negocio en constante adaptación:

Actualmente las empresas se enfrentan a un mercado que evoluciona más rápidamente que nunca. Productos y servicios así como las formas de producirlos, venderlos y promocionarlos quedan anticuados y obsoletos en una fracción del tiempo en que ocurría diez o veinte años atrás. En esta carrera atroz por conservar las ventajas competitivas y obtener algunas nuevas, el aporte de conocimiento que nos da el análisis de toda la información descrita en los puntos anteriores se antoja primordial. Cualquier ventaja nos separa del éxito o el fracaso. En estos momentos, más que nunca, es absolutamente crucial obtener valor de toda la información disponible para mejorar nuestra operación y nuestra estrategia

En conclusión, parece que a la vista de todo lo anterior, sí que estamos ante un momento en el que quedarse fuera de la utilización de BIG DATA puede significar perder un tren importante hacia una nueva forma de entender tanto el mercado en que competimos como nuestro propio desempeño. Tiene pinta de ser más de lo que cualquier empresa se puede permitir.

¿Y cómo le hincamos el diente?

La única forma de aproximarnos con éxito al desafío que Big Data nos presenta es ir avanzando por etapas y creciendo a partir de un núcleo inicial formando una espiral. Las fases podrían ser las siguientes:

1. Construir un EDW para aprovechar analíticamente la información estructurada interna existente en la empresa.
2. Enriquecer este EDW con información externa comprada a terceros
3. Aprovechar la información no estructurada interna proveniente de nuestra Web, Contact Center o canales en Redes Sociales, para ello construir un sistema de almacenamiento no estructurado que se conecte con el EDW.
4. Mejorar la organización, gestión y monitorización de la red de actividades de la empresa con sistemas BPM y BAM, e integrar la información conseguida en los sistemas anteriores.
5. Por último capturar información desestructurada externa de la Social Web, sólo aquella que tenga valor (la cuarta "V") para nuestro negocio, almacenarla en sistemas específicos para información no estructurada y procesarla para obtener de ella el máximo provecho mediante análisis semántico, clasificaciones heurísticas, etc.

En cada una de estas fases avanzaríamos con una metodología clásica de Inteligencia de negocio, es decir:

1. **DATOS:** Construcción del modelo de datos a utilizar, seleccionar la información fuente, preparar procesos de calidad de datos, procesos de extracción, transformación y carga, crear los metadatos que den sentido de negocio a todos los valores, ajustar las ventanas de carga, etc
2. **INFORMACION:** Definir y crear los Key Performance Indicators que mejor nos represente el estado de los distintos aspectos de los problemas de negocio a estudiar: indicadores financieros, de cliente, de ventas , de producción,... Definir y creas los cuadros de mando y sistemas de reporting que hagan visible estos KPI de la forma más útil posible para quien lo utilice.
3. **CONOCIMIENTO:** Buscar información oculta mediante el uso de Minería de Datos (estadística y matemática avanzada), crear nuevos KPI que me den una visión sobre las posibilidades futuras: Estudios de previsión de demanda, de abandono de cliente,

de posible fraude, de cuál sería nuestra mejor próxima oferta, de propensión a la compra,...

4. **ACCIÓN:** Una vez tengamos un enorme arsenal de información y conocimiento disponible podemos optimizar las Redes de Actividades de la empresa, automatizando la toma de decisiones operacional, y ayudando de forma decisiva en la toma de decisiones estratégicas. Obviamente, hay que conseguir una Estrategia que nos lleve al éxito, y una Red de Actividades que cumpla la Estrategia.

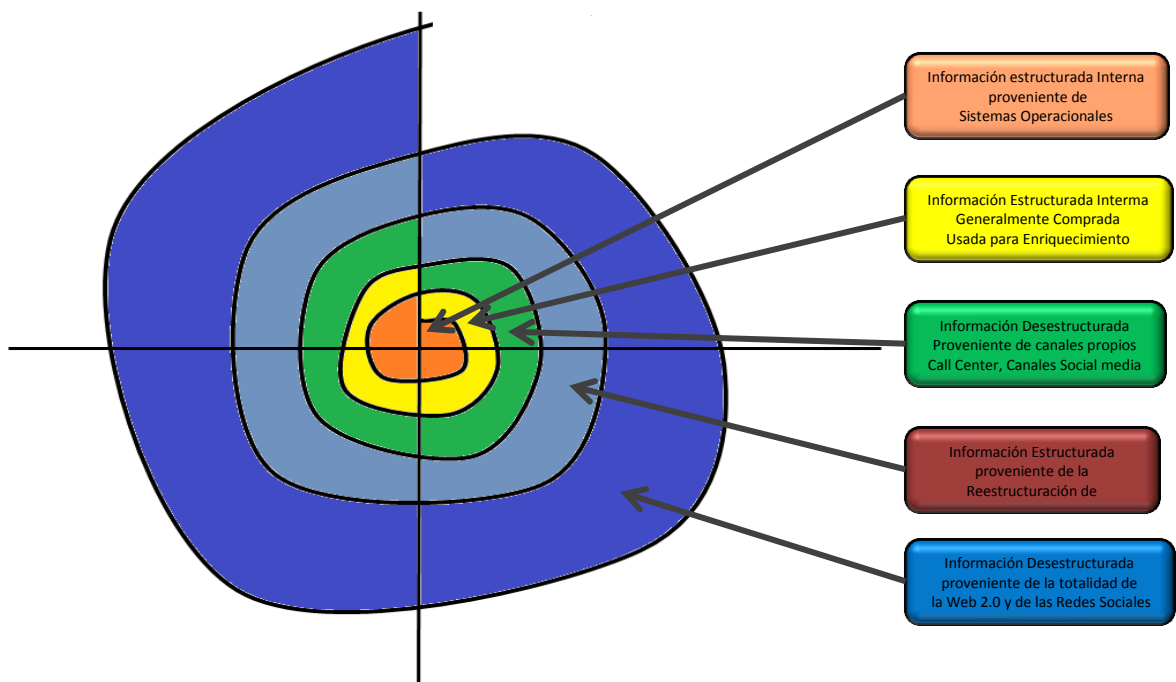


Imagen del autor junto con el cuadro del tinerfeño Vicente López

Evidentemente es un proceso complejo, que involucra múltiples tecnologías y metodologías dependiendo de la naturaleza de los problemas a tratar y de la información que queramos utilizar en cada nueva iteración de la espiral, pero para nada imposible. Además, el premio al esfuerzo de llevarlo a cabo es realmente valioso: conseguir para nuestra empresa una posición de ventaja muy importante al manejar información de que los otros competidores no disponen y ponerla en acción tanto en la estrategia como en la operación.

El cuadro que inspira el logo de Habber Tec, “Jugada Maestra” del tinerfeño Vicente López, ilustra perfectamente este escenario: Una empresa

representada por un caballo de ajedrez, pintado a colores, enriquecido con la enorme potencia que le otorga el disponer de toda la información relevante de si mismo y de su entorno, capaz de desarrollar y ejecutar por tanto una estrategia superior, dominando y derrotando al resto de competidores, piezas blancas y negras que comparten el tablero, incapaces de evolucionar a su mismo ritmo ni de cambiar las reglas del juego.

Con la colaboración de:





*Confederación Española de
Directivos y Ejecutivos*

World Trade Center

Moll de Barcelona, s/n Edificio Este 1ª planta
08039 Barcelona (Spain)

Tel. +34 93 508 83 20
Fax. +34 93 508 83 21

Correo: info@directivoscede.com

Acceda a nuestra web:

www.directivoscede.com

La Confederación Española de Directivos y Ejecutivos (CEDE), fundada por la Asociación Española de Directivos (AED) y la Asociación Española de Ejecutivos de Finanzas (AEEF), tiene como objetivo principal agrupar a distintas asociaciones de directivos al objeto de dotarlas de una representación en consonancia con la relevante función que desempeñan en el ámbito socioeconómico, así como estimular sus actividades y coordinar acciones en beneficio de las entidades confederadas y de todos sus socios individuales. Desde su fundación en 1997, se han integrado en la Confederación otras asociaciones de directivos pertenecientes a diversos sectores y ámbitos territoriales. En la actualidad CEDE agrupa 46 entidades y más de 120.000 directivos, por lo que está alcanzando una representatividad significativa, que continuará incrementándose en el futuro con la incorporación de nuevas entidades.

Sobre la Comisión de Gestión del conocimiento empresarial

Si aceptamos que hoy en día el conocimiento es un factor esencial para el desarrollo empresarial y que los activos intangibles son, cada vez más, la parte creciente del valor y de la capacidad de generar resultados sostenibles, se considera de especial relevancia incorporar en el programa general de actividades de CEDE la actuación de un Observatorio sobre Gestión del Conocimiento Empresarial, Capital Intelectual e Innovación, que será gestionado por esta comisión de trabajo.

La comisión de trabajo considera prioritario enlazar cualquier iniciativa que encaje con la mejora de la gestión empresarial y, específicamente, con el incremento de la calidad de las funciones directivas y ejecutivas. Por tanto, el enfoque de sus actividades será eminentemente práctico, obviando perspectivas excesivamente académicas.

También sus actuaciones se enmarcan en un contexto global, alineadas con acciones y trabajos realizados en el marco de la Unión Europea.

En cuanto a los objetivos de trabajo, principalmente son:

a) Seguimiento de las principales corrientes de opinión y avances teórico-prácticos en materia de activos intangibles, capital intelectual, gestión de la información, inteligencia empresarial y tecnologías de gestión del conocimiento.

b) Acuerdos con fundaciones y entidades de estudio nacionales e internacionales, cuyo ámbito de investigación merezca ser seguido, conocido y divulgado por CEDE.

c) Difusión en los medios virtuales de CEDE de las mejores prácticas y los avances más significativos en Gestión del Conocimiento, Capital Intelectual e Innovación.

d) Creación de plataformas de debate y aprendizaje, de modo que pueda mantenerse una Red de intercambio de Conocimiento entre las asociaciones de CEDE.

Puede conocer más sobre nuestra actividades y cuadernos en la web de CEDE o solicitando información en el email info@directivoscede.com