

# “Big Data”

## O CÓMO EXTRAER TODO EL VALOR DE “TODA” LA INFORMACIÓN

En 2011 el volumen de datos almacenados alcanzó los 1,8 zettabytes, 1,8 millones de Petabytes o 1,8 billones de Gigabytes. Buscando alguna analogía que permita entender la magnitud del fenómeno, 1,8 ZB, equivaldrían a 57,5 mil millones de iPads de 32 GB, suficiente para reconstruir la Gran Muralla de China al doble de su altura. Para el final de esta década, se espera que esta cifra aumente hasta 35 zettabytes (1). Vivimos en un mundo que crea y replica información a un ritmo vertiginoso, y estas masas de datos no tienen la forma tradicional de información estructurada en grandes bases de datos, sino que hoy invariablemente llegan en enormes masas de datos no estructurados, conocidos como “Big Data”.

POR JOAQUIN POTEL

Este gran volumen de información afecta por igual a los organismos públicos y empresas privadas, y para ambas supone un reto gestionar estos crecimientos complejos de información en una situación como la actual, en la que los presupuestos se reducen casia diario. Existe un obvio desfase entre la inversión en tecnologías de la información y el crecimiento de la misma, al menos en cuanto a volumen a gestionar se refiere.

¿Pero de dónde procede toda esta información? La respuesta está en la proliferación de las comunicaciones digitales. En gran parte, mucha de la comunicación que tradicionalmente ocurría de forma analógica se está digitalizando a gran escala, y por otro lado, el volumen de comunicación se ha disparado y globalizado. En 2009 había 450 millones de usuarios de internet móvil. En 2013 se prevé que esta cifra aumente a más de mil millones (2). Los medios sociales generan enormes cantidades de datos no estructurados a medida que proliferan las formas de comunicación vir-

tuales. Nuestro mundo está lleno de sensores inteligentes con capacidad para recibir información permanentemente y en casi cualquier lugar, que puede ser tratada por entornos M2M (Máquina a Máquina). Las aplicaciones de software que se pueden descargar permiten a las personas hacer casi cualquier cosa desde sus teléfonos móviles y tabletas, mientras que

las aplicaciones geoespaciales como Google Maps generan enormes cantidades de datos todos los días. A todo ello hay que añadir el consumo masivo de fotografía y video digital todos los días. Un auténtico universo digital.

A este fenómeno no escapan las administraciones públicas. Los ciudadanos requieren poder solucionar



sus gestiones con los organismos públicos empleando los mismos medios digitales con los que se comunican con su banco o sus amigos. Por otra parte, tanto a nivel nacional como supranacional, cada vez hay más requisitos legales sobre como tratar y gestionar esta información. De hecho, las administraciones públicas están destinadas en cierta medida a liderar la gestión de este diluvio de información digital.

El solo hecho de mantenerse al día con los requisitos de procesamiento y almacenamiento de este tsunami de datos es una tarea compleja en sí misma. Pero extraer una información significativa de cantidades colosales de datos es un reto totalmente diferente. Y es aquí donde se encuentra el verdadero valor, tanto para las empresas como para los organismos públicos. Las tecnologías de la información pueden generar valor añadido abordando planteamientos y formas de hacer que hasta hace nada eran impensables.

### **Gestionar datos de gran tamaño**

EMC está respondiendo al reto que representa Big Data con productos innovadores de hardware y software. Entre estos hay nuevas arquitecturas de almacenamiento de escalabilidad horizontal (scale-out) diseñadas específicamente para almacenar y gestionar volúmenes únicos de información no estructurada que escalan de forma lineal hasta decenas de Petabytes.

Aunque el informe Digital Universal de 2011 anticipa que solo el 2% del gasto de TI de este año se dedicará a servicios en la nube, se espera que este porcentaje suba exponencialmente hasta un 20% en 2015 (3). Más de un tercio de empresas dicen que ya utilizan una mezcla de

infraestructuras en la nube pública y privada (4). Las tecnologías de nube (entornos virtualizados sobre arquitecturas x86 y que escalan sin límite de forma lineal) ofrecen muchas ventajas al consolidar la gestión y administración, optimizar el consumo de recursos, acelerar el aprovisionamiento de infraestructura y aprovechar economías de escala. Estas arquitecturas de nube suponen un facilitador clave para poder procesar entornos de Big Data.

### **Extraer valor del caos**

No es solo la cantidad de datos, sino la naturaleza de los mismos lo que está cambiando. Están muy poco estructurados, se replican a una velocidad increíble y gran parte de los mismos se disuelve con la misma rapidez con la que se materializa. Hay volúmenes ingentes de datos de los que tan sólo quiero extraer la información relevante en el momento adecuado. Los Big Data requieren un tratamiento diferente y una tecnología mucho más poderosa que antes para gestionarlos y manipularlos.

De modo que, ¿dónde deberíamos empezar cuando se trata de extraer valor del universo digital en el entorno de la administración pública? Primero identificando las ventajas para el usuario-ciudadano de invertir en la tecnología necesaria para automatizar el proceso de captura, procesamiento y almacenamiento de datos. A continuación, planificando una estrategia de inteligencia de datos evolutiva, que capacite a cada organismo para entender e interpretar los datos que maneja, dando como resultado beneficios tangibles para el servicio que presta tanto de forma interna como para los ciudadanos.

El análisis de los Big Data permite a las instituciones cruzar y compara

rar conjuntos dispares de datos sin estructurar, para con ello encontrar nueva información capaz de prestar un mejor servicio con un menor coste. Por ejemplo, los organismos responsables de hacer cumplir la ley podrían detener a muchos más sospechosos si analizaran y compararan el contenido de las televisiones de circuito cerrado, el vídeo, las evidencias fotográficas, los registros de la policía, los comentarios de los medios sociales y otras fuentes de información. La posibilidad de cruzar datos de diferentes organismos también aportaría grandes beneficios en la lucha contra el fraude fiscal.

En la sanidad, por ejemplo, los Estados Unidos emplean los Big Data para detectar qué ciudadanos emplean servicios de urgencia de forma continuada, que patologías padecen, y aplicar la medicina preventiva.

Para aplicar la metodología de los Big Data son fundamentales dos consideraciones. La primera, decidir qué información será la más valiosa para la labor que desempeñan nuestras instituciones. La segunda, y tal vez más importante, es que el análisis de los Big Data requiere de una interpretación sofisticada y de una serie de habilidades completa y multidisciplinaria para obtener resultados. 📊

### **NOTAS**

1. IDC iView 'Extracting Value from Chaos', junio de 2011, John Gantz y David Reinsel.
2. IDC, John Gantz, diciembre de 2010
3. IDC, op. cit.
4. Investigación IDC-Unisys, julio de 2011

Joaquín Potel  
Director Comercial de Grandes Cuentas  
de EMC