

Ciclo de Desayunos Fundación ASTIC 2012

Big Data: Retos y oportunidades para las Administraciones Públicas

POR MAOLE CEREZO
REDACTORA JEFE DE BOLETIC

FOTO AITOR DIAGO

Evento patrocinado por



En el desayuno que celebramos el pasado mes de septiembre se reflexionó sobre Big Data. Ana Bajo, Vicepresidenta de ASTIC, abrió el evento abordando el Big Data como un concepto que surge por una evolución natural dentro de la era de la información, en una fase de la misma caracterizada por los grandes y continuos incrementos de almacenamiento global de datos y cuyos rasgos definitorios vienen determinados por lo que se ha venido en llamar las “tres uves”: volumen, velocidad y variedad. Si bien, en la Administración, el almacenamiento supone un gran coste, “también genera sinergias como la reutilización de información o las políticas de Open Data”.

La evolución de Big Data es tan rápida, señaló, que en “la actualidad,

la demanda de perfiles profesionales apenas alcanza a cubrir el 60% de la oferta”. Los directivos deben actualizarse y tomar conciencia de la importancia real del tratamiento automatizado de los datos en los procesos de toma de decisiones.

Antes de abordar cualquier proyecto relacionado con Big Data “es necesario identificar los objetivos, seleccionar las fuentes de información, y evaluar las técnicas de análisis adecuadas” puntualizó Bajo. También será una cuestión primordial “determinar el perfil del responsable del proyecto como aglutinador de la diversidad de fuentes y de colaboradores”.

La Vicepresidenta cedió la palabra a Alejandro Giménez, Chief Technology Officer de EMC España España para quien Big Data “no nace en el seno de

las TIC, sino que surge de la propia sociedad, fruto de los cambios sociales, de la “vida digital”, que es la determinante de las “tres uves”: volumen ingente de datos, gran variedad de fuentes y formatos y velocidad, tanto de generación de información como de su tratamiento”.

Cada vez, en mayor medida, “se intenta hacer un uso predictivo de la información, y de esta forma aparece la “cuarta uve”, el valor, que se extrae mediante la aplicación de las técnicas que conforman el Big Data”. Porque, en general, los sistemas “no están preparados para afrontar el Big Data, término que puede representar conceptos distintos dependiendo de cada interlocutor, pero que está transformando el servicio y el negocio”, señaló el directivo. »



Respecto a las tecnologías clave en Big Data, según Giménez, “el cloud computing desempeña un importante papel como habilitador o catalizador, facilitando el desarrollo de proyectos y proporcionando acceso a datos para los mismos desde diferentes lugares. Pero otras tecnologías cobrarán relevancia de forma progresiva. Es el caso de los “procesamientos masivamente paralelos”: cada vez se almacenan más datos, y “se persigue trabajar con ellos en conjunto, desde un punto de vista global, cuyo procesamiento tiene que evolucionar hacia la utilización de programas o tratamientos analíticos capaces de correr en muchos sistemas paralelamente”. También cobrarán importancia los “algoritmos paralelos”: aparecen grandes volúmenes de información no estructurada que queremos analizar, lo que requiere de adaptaciones sobre algoritmos y analíticas.

En relación con los “procesa-

mientos masivamente paralelos” y el almacenamiento de la información, el directivo de EMC abordó la cuestión del “escalado lineal: los sistemas deben requerir cada vez más recursos, y se hace necesario predecir con fiabilidad su rendimiento en función del crecimiento que afronten. Por su parte, la automatización es cada vez más requerida, y cobrará importancia de forma progresiva para tareas como el almacenamiento de la información, su distribución, su análisis y, en definitiva, para el tratamiento adecuado de la información. Adicionalmente, cuestiones relativas a la seguridad, o de naturaleza ética evolucionarán a la estela de la tecnología, y éste hecho afectará a los desarrollos normativos que se acometan”, apuntó Giménez.

Business Intelligence/Big Data

¿Y dónde se puede situar la línea divisoria entre Business Intelligence y Big

Data? Le preguntó Ana Bajo.

A lo que Alejandro Giménez contestó que existen muchas percepciones del concepto de Big Data, y esa línea suele ser difusa. “La obtención de datos de múltiples fuentes, sean estructurados o desestructurados, el cruce reiterado de esa información entre sí y con información externa, y su análisis para extraer inteligencia es Big Data”.

José Manuel Pacho, de la Consejería de Sanidad de la Comunidad de Madrid, puso de manifiesto la importancia del “valor” como “cuarta uve” ya que, a su juicio, “es el elemento que siempre habrá que tener en cuenta a la hora de evaluar la posibilidad de afrontar un proyecto de Big Data”. A la vez, se interesó por conocer algún proyecto basado en Big Data como generador de eficiencia

En el campo de las Telecomunicaciones, le respondió el CTO de EMC

para España, “la competencia se basa en combinar ancho de banda y precio buscando alcanzar la ecuación ideal. En función del tráfico, se han detectado datos de consumo y usuarios que aportan eficiencia al contribuir, de forma decisiva, a optimizar dicha ecuación, evitando así intensificar el trabajo de expansión de la oferta”.

Guillermo Searle, de la Agencia Tributaria, planteó la cuestión de los sistemas de medición de información, de conocimiento, etc., aplicables en el campo del Big Data a la hora de dimensionar cuantitativamente su aplicación. Para EMC, “a la hora de evaluar una estrategia Big Data, se pueden aplicar técnicas de escalado lineal y automatización, que permiten afrontar almacenamientos gigantescos sin tener que afrontar migraciones de datos que serían cada vez más complejas”. En el plano meramente cuantitativo, “se ha superado con alguno de nuestros clientes el millón de Terabytes de información”.

Para Angel Luis Sánchez, del Servicio Madrileño de Salud, “hay que tener visión para encontrar el valor”. En el caso concreto de la Sanidad, señaló que “los estudios genéticos o la ayuda al autodiagnóstico serán campos de aplicación de Big Data, pero habrá que reducir previamente su coste de forma drástica, ya que será necesaria mucha inversión en investigación para ser capaces de extraer información de las historias clínicas”.

Como ejemplo de obtención de grandes volúmenes de información a través de su propia actividad, para su cruce con fuentes públicas y posterior extracción de nueva información útil para las propias empresas, el CTO de EMC se refirió a las compañías de seguros.

Agencia Tributaria

Un caso concreto de la AAPP que sacó a colación Román Díez, de la Agencia Tributaria, fue el de su organismo. Explicó como abordan el trabajo los profesionales TIC destinados dentro de la Agencia para la lucha contra el fraude. Para esta tarea “disponen de datos estructurados procedentes de fuentes reconocidas y de datos poco estructurados extraídos de Internet y cuya utilización acarrea gran dificultad”. ¿Podrían incorporarse estos últimos a las herramientas de análisis sin la necesidad de su almacenamiento? ¿Qué posibilidades existen de integración de estos datos con los internos en el momento del análisis sin que sea necesario su almacenamiento estable? Preguntó.

Alejandro Giménez hizo referencia a las distintas tecnologías que permiten el análisis de la información no estructurada, así como a las herramientas que analizan como servicio grandes masas de información aplicando un coste según unidad de volumen. La información “se puede analizar “en directo”, sin necesidad de almacenamiento, si bien es necesario tener muy claro que no se requerirán nuevos tratamientos de esos mismos datos”.

El problema que supone el rápido incremento de los datos almacenados y el coste de mantenimiento de los mismos, muchas veces de poca utilidad, lo cual no solo encarece su almacenamiento, sino que complica su tratamiento fue un tema que pusieron sobre la mesa Rocío Montalbán, del Ministerio de Industria y Manuel Alonso, de la Intervención General.

La actuación de EMC se basa sobre todo, como explicó Giménez, “en tres aspectos del Big Data: el almacenamiento basado en el escalado lineal, que permite acometer un crecimiento

modular; la capacidad analítica, a través de bases de datos propias con procesamiento masivamente paralelo y escalado lineal (“scale out”) capaces de obtener y analizar datos procedentes de bases de datos externas y, por último, una gestión documental orientada al usuario”.

La última intervención de los asistentes de la Administración corrió a cargo de Ignacio Bellido, de la Agencia de Informática y Comunicaciones de la Comunidad de Madrid, para quien la información en poder de la Administración “tiene unos usos muy concretos y una difusión muy restringida, al amparo de la legislación relativa a protección de datos”. Por ello, preguntó por el impacto que puede generar ese marco regulatorio en la implantación de un proyecto Big Data.

Alejandro Giménez explicó que puede darse la situación que “el cruce reiterado de de datos procedentes de fuentes públicas acabe generando la publicación de datos que puedan ser considerados privados”. En todo caso, “la preocupación por la seguridad de los datos es creciente en el seno de todas las organizaciones. Está previsto desarrollar auditorías sobre la información residente dentro del Big Data, que permitan expurgar aquellos datos que incumplen la legislación”. De cualquier forma, se trata “de analizar las reglamentaciones que afectan al proyecto y evaluar, en todo momento, el grado de cumplimiento de esa normativa”. 