

Suplemento de la revista Computing 690 - 1 julio 2012

computing

ESPAÑA

Big Data



En la Era del Big Data

EMC²

Las bases de datos tradicionales no son las más indicadas para soportar el Big Data

Gestionar grandes datos requiere una nueva aproximación en tecnología y profesionales

El volumen cada vez mayor y el detalle de la información capturada por las empresas, el aumento de las redes sociales y el fenómeno de Internet de las cosas están impulsando el crecimiento exponencial de los datos a un ritmo

trepidante. Lidar con el manejo de estos Big Data para aprovecharlos y analizarlos de cara a una mejor toma de

decisiones requiere evolucionar el enfoque en términos de almacenamiento, aplicaciones de analítica y administradores de las bases de datos.



● La rápida explosión de los datos no estructurados, catalizados por la omnipresencia de Internet y el masivo crecimiento de los dispositivos móviles, ha creado un nuevo mercado para capturar y analizar datos que no son tradicionales. Ahora, las fuentes de información son sensores móviles, redes sociales, medios de pago electrónicos, sistemas de videovigilancia, vídeos, redes inteligentes, imágenes médicas... Estamos hablando del fenómeno Big Data, y los primeros afectados de esta tendencia son los proveedores de bases de datos, pues deben apoyarse en nuevas tecnologías que ayuden a sus clientes a lidiar con tanta cantidad de información.

La cuestión es que Big Data no está correctamente soportado por las bases de datos tradicionales debido a que ahora la

mayoría de los datos son desestructurados y demasiado voluminosos, lo que conlleva unos costes prohibitivos en hardware y software. IDC, por ejemplo, señala que en esta década el universo digital se multiplicará por 44 veces, pasando

Hay que evolucionar e ir más allá del BI tradicional, y colocar a los analistas de datos en el centro de toda decisión

de 0,9 a 35,2 zetabytes; siendo el 90% de dicha información datos no estructurados.

Para capitalizar el Big Data, especialistas como EMC, están combinando hardware, código abierto y tecnología de última

generación para capturar y analizar dicha información. Un análisis que permita a las empresas tomar decisiones más acertadas al adquirir nuevas perspectivas derivadas de los datos empresariales. Por este motivo, el Big Data es una

oportunidad enorme para aquellas compañías, que entienden que deben evolucionar más allá del Business Intelligence tradicional colocando a analistas de datos en el centro de las decisiones de negocio.

Este movimiento permite, entre muchas otras cosas, mejorar el margen de negocio al poder alinear la oferta con las necesidades del cliente, utilizar recomendaciones orientadas a beneficios, o identificar a clientes en riesgo.

El camino a Big Data

Tener acceso a todos los datos para obtener respuestas más rápidas requiere de sistemas elásticos y escalables. Para ello, EMC sugiere evolucionar el enfoque y optar por infraestructuras orientadas al Big Data, tanto en los almacenamientos incluyendo tecnologías cloud, como en los motores analíticos orientados a información estructurada o desestructurada. Así, tecnologías como Isilon de EMC permiten ofrecer un repositorio único de más de 10 petabytes

¿Por qué el Big Data es mágico?

Decir que el Big Data no tiene relevancia para las organizaciones es como si se afirmara en los años 90 que Internet no es importante para los negocios. Teniendo en cuenta tres fuentes de datos clave, el Big Data sí es primordial.

La clave para gestionar el Big Data es cambiar la relación entre el departamento de TI y el negocio, pues el verdadero desafío es comprender los fundamentos de lo que la empresa hace y encontrar los modelos de análisis de datos que permitan crear valor adicional para el negocio.

Tecnología Operacional: Hace referencia al Internet de las cosas, la instrumentación de casi todo. Si la empresa trabaja con materiales físicos, debería considerar cómo una información completa sobre dicho material podría cambiar su operativa, costes y eficiencia. Por ejemplo, la industria de seguros de automóviles siempre se ha basado en el riesgo medio derivado de la información demográfica básica. Hoy en día, está empleando sensores en los vehículos para evaluar el riesgo de cada individuo en función de sus hábitos de conducción. Este cambio fundamental en la forma en cómo se calculan los riesgos permite reducir las primas y aumentar los beneficios.

Datos de Terceros: Se trata de información que está disponible si se va en busca de ella. Tanto las administraciones como empresas de todo tipo están abriendo el acceso a sus datos para el consumo externo. Si se combinan los datos ya disponibles con sus nuevas fuentes, podría proporcionar una nueva perspectiva a la toma de decisiones. Por ejemplo, las proyecciones demográficas de los datos del censo podrían ser muy valiosas para decidir sobre los desarrollos de infraestructura, la ubicación de nuevos almacenes, puestos de torres de telecomunicaciones, etc.

Redes Sociales: Si la empresa vende productos o servicios tiene que aprovechar la mina de oro de las redes sociales. Atrás han quedado los días en los que era posible controlar uno mismo la marca. La creciente influencia de las redes sociales y la voz de la multitud se convertirán en la principal fuerza motriz para la percepción de la marca.

en lugar de islas de datos, que suelen requerir una administración manual y estar soportadas en sistemas dispares. Además, los sistemas tradicionales de almacenamiento suelen ser un servicio provisionado por el departamento de TI, con el consiguiente consumo de recursos internos. En cambio, con la tecnología Isilon, se tiene un sistema único geo-distribuido, la gestión se realiza mediante políticas automáticas y el enfoque es de autoservicio.

Resumiendo, guardar la información en sistemas idóneos para Big Data permite realizar una gestión más eficiente, facilita una escalabilidad más transparente; proporciona una automatización de situación, protección y servicios del dato; simplifica el acceso a lo largo de redes y plataformas; y optimiza la medición y autoservicio entre los sistemas usuarios.

EMC recomienda emplear también herramientas ágiles de analítica, con arquitecturas MPP (Procesamiento masiva-

mente paralelo) focalizadas en escalabilidad y productividad, así como reforzar la colaboración en tiempo real y el auto-servicio. Greenplum es la opción de base de datos orientada a analíticas de EMC.

En cuanto al análisis de datos no estructurados, las soluciones tradicionales suelen ser reactivas, predefinidas y lentas proporcionando una información limitada, y analizando sólo terabytes de datos estructurados y agrupados. En cambio, las aplicaciones diseñadas específicamente para Big Data ofrecen una analítica más ágil, proactiva, de información expandida y correlacionada, y optimizan el cumplimiento normativo. Son capaces de analizar hasta PB de datos crudos, tanto estructurados como desestructurados; y son compatibles con los sistemas y herramientas existentes al estar abiertas a un amplio ecosistema de soluciones. En este sentido, EMC ofrece la integración entre su base de datos Greenplum y Hadoop, como solución orientada al análisis de datos desestructurados.

En los procesos analíticos tradicionales, los administradores de bases de datos pueden convertirse en un cuello de botella para los equipos de Business Intelligence. Por tanto, la organización necesita evolucionar hacia esquemas más colaborativos formados por los administradores de bases de datos, adscritos al departamento de TI, y con equipos dedicados a la ciencia de los datos, que incluirían al administrador de la plataforma de datos, usuarios de Business Intelligence, usuarios de la línea de negocio, y un nuevo perfil de profesional: los científicos de datos. Precisamente este sería el siguiente paso en el camino al Big Data.

Con el objetivo de aportar un nuevo valor al negocio potenciando las aplicaciones EMC propone emplear soluciones que permitan tomar decisiones en tiempo real, monetizando la información. Frente a unas analíticas que ofrecían datos agrupados del pasado, estimativas y aisladas de las aplicaciones, ahora se impone proporcionar la información en tiempo real, junto con una orientación predictiva y embebida en las aplicaciones.

En conclusión, la idea de EMC es poner la información en acción de una manera sencilla con herramientas para no

Forrester Wave: Mercado de proveedores de soluciones empresariales Hadoop, 1Q2012



programadores y con modelización de procesos, que ayudan a tomar decisiones, in-

El reto es que las empresas ahora quieren saber mucho más de lo que los enfoques

redujeron en una región determinada, pero no sabría decir el porqué. Precisamente, son estos 'porqués' los más útiles para las compañías y que los análisis avanzados de datos saben responder. Explicar el pasado y predecir el futuro. Así, si la empresa mencionada anteriormente descubre que la presión competitiva en los precios es la culpable de la disminución de las ventas, podría explorar qué sucedería si cambia su estrategia de marketing y hacer predicciones precisas acerca de si esto podría ayudar a impulsar las ventas del

producto X en las regiones donde las ventas han bajado.

Obviamente, hay mucha información y diversos métodos de modelización. Los análisis de hoy en día requieren el uso de técnicas estadísticas para 'aprender' de los datos, de patrones de relieve y anomalías, y de profesionales que sepan utilizarlas. El empleo de tecnologías Big Data no solo permite aumentar la capacidad de procesamiento, también se trata de encontrar esas ideas que marcan la diferencia, siempre y cuando se disponga de los perfiles y experiencia para llevarlo a cabo.

Es aquí donde surge el científico de datos, un término que se aplica a aquellos profesionales, con habilidades en matemáticas, estadísticas e ingeniería informática, que son capaces de extraer el máximo valor de los datos de la organización, cerrando la brecha entre las necesidades del negocio y las TI. Sin embargo, es crucial que la comunicación entre ellos y la organización sea lo más fluida posible. Contar con plataformas y comunidades de desarrolladores y científicos de datos es una buena opción para fomentar esta colaboración y la innovación.

Contar con plataformas y comunidades de científicos de datos y desarrolladores fomenta la innovación y colaboración

cluso de contexto, más rápidas y mejores en función del último dato recabado; y que faciliten un seguimiento de la información para reducir los riesgos asociados a los datos.

tradicionales de análisis de datos les ofrecen. Por ejemplo, estas analíticas pueden decir a una organización que las ventas del producto X se

El científico de los datos

Conviene destacar que la analítica avanzada no necesariamente tiene que estar relacionada con el Big Data; es igualmente eficaz para mejorar la forma en que las empresas pueden obtener información valiosa del Business Intelligence en sus bases de datos, ya sean grandes o pequeñas. Estos datos por lo general incluyen cifras e información de áreas como ventas, demografía de marketing, registros de CRM y otras funciones básicas del negocio. Estos son los datos estructurados que las organizaciones saben utilizar bien.



Una excelente oportunidad para organizaciones de cualquier ámbito

Big Data es una gran fuerza de cambio

Big Data ya es una realidad, y se está convirtiendo en una fuerza formidable para el cambio, como lo demuestran los proyectos, desarrollados por EMC, y que se muestran en

este reportaje. La gestión y análisis de los grandes datos, estructurados y no estructurados, aplicados en campos como la investigación científica, sanidad, seguridad, re-

des sociales o medios de comunicación, entre otros, constituye para las empresas una herramienta única de ganar competitividad y de mejora de la vida ciudadana.

Desarrollar productos sociales innovadores

LinkedIn es una de las empresas con una visión audaz para el futuro del Big Data y la web. De hecho, el fundador de la compañía, Reid Hoffman, ha declarado que la Web 3.0 procederá de los datos y de la analítica. Con más de 100 millones de usuarios, LinkedIn está en una posición privilegiada para acceder a sus datos y aprovechar sus puntos de vista. El compromiso de la compañía para la búsqueda de aplicaciones innovadoras y útiles se puede encontrar en productos como LinkedIn Skills, listas con hipervínculos de habilidades que permiten saber, por ejemplo, cuáles son las habilidades particulares más interesantes para un posible empleador. Así, los usuarios pueden encontrar personas, empresas y puestos de trabajo relacionados con sus habilidades. Hay que mencionar también un proyecto de Big Data en LinkedIn, como es InMaps, una visualización de datos innovadora que cartografía

las conexiones en todo el universo profesional de un usuario.

En definitiva, LinkedIn aprovecha los datos de alto valor generados por los profesionales que han establecido sus identidades reales en este servicio. Al disponer de un grupo científico de datos, -que incluye a científicos de diversas procedencias y profesionales con experiencia en visualización-, LinkedIn ha proporcionado constantemente a su base de usuarios formas innovadoras de conexión y de construir su reputación profesional. Este grupo ofrece a LinkedIn un enfoque único y competitivo, permitiendo a la compañía ofrecer herramientas profesionales antes que cualquier otra empresa en el espacio social. Y "esto es solo el inicio de la revolución Big Data", declara Peter Skomoroch, científico principal de datos en LinkedIn.

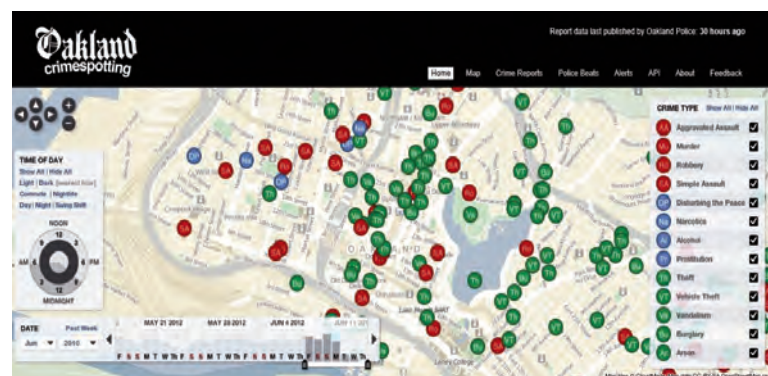
Transformar la investigación genética con tecnología de escala

El Instituto Broad, un centro de investigación de medicina genómica formado por científicos del Instituto de Tecnología de Massachusetts (MIT) y de la Universidad de Harvard, quería llegar a una comprensión global de los elementos en el genoma humano e identificar cómo los circuitos de las células genómicas procesan la información. A través de una serie de proyectos de colaboración, el Instituto está identificando las mutaciones celulares que causan el cáncer, la base molecular de los virus, bacterias y otros agentes patógenos que causan enfermedades infecciosas; e innovando en el desarrollo de fármacos. Todo ello ha contribuido a incrementar exponencialmente los datos que maneja el Broad Institute, -el año pasado creció a ocho petabytes-, con un rendimiento que se duplica cada cinco meses. Para gestionar los cada vez más grandes conjuntos de datos relacionados con el genoma, los investigadores del Instituto Broad se basan en el almacenamiento en cluster de EMC Isilon, y han desarrollado más de 30 herramientas de software a medida para realizar un análisis altamente especializado de dicha información. En línea con su compromiso con la investigación y colaboración abierta, el software y los datos están disponibles para todos los investigadores para su descarga desde el sitio web del Broad Institute.



Big Data da nuevos conocimientos sobre la vida urbana

En la era del Big Data, tener en cuenta fuentes de datos innovadoras en las ciudades puede revelar nuevos conocimientos sobre la vida urbana. Es el caso de los proyectos Cabspotting y Crimespotting, desarrollados por la firma norteamericana Stamen Design, que permiten visualizar tendencias económicas, sociales y culturales nunca antes vistas. Crimespotting nació porque las aplicaciones de cartografía del crimen, utilizadas por la Policía de Oakland, tenían una funcionalidad limitada y mapas anticuados. Todos los días, Crimespotting recoge las denuncias de delitos del servicio Crimewatch de la ciudad y los importa a su base de datos. Luego ofrece un entorno cartográfico interactivo que permite seguir los casos de delincuencia en el vecindario, y proporciona servicios personalizados como alertas específicas para la vecindad. Cabspotting es otro de los proyectos de la empresa, que utiliza las coordenadas GPS de los taxis de San Francisco para revelar patrones ocultos de la vida urbana. Rastrea las carreras de los taxis situando sus movimientos en un mapa interactivo de la ciudad. La información se recoge desde un servidor de la sede central de Yellow Cab, que recaba los datos GPS de cada taxi en tiempo real, y una vez analizados, el resultado es un nuevo tipo de antropología urbana que habla de asuntos políticos, sociales y económicos.



Big Data, en la línea de defensa contra las pandemias



El mundo está conectado a una escala sin precedentes, y el riesgo de amenaza de una pandemia mundial ha aumentado de manera exponencial. Tradicionalmente, los esfuerzos de control de enfermedades reaccionan cuando surgen los brotes. Sin embargo, estos tiempos de respuesta retardada son demasiado lentos para una población que está siempre en movimiento. Además, los virus son cada vez más resistentes a los antibióticos, lo que hace ineficaces las medidas preventivas tradicionales. Es necesario un nuevo enfoque para el control de enfermedades.

En esta línea, Previsión Global Viral (GVF) aprovecha el Big Data para evitar grandes pandemias mundiales antes de que comiencen, ante "fracasos como la continua falta de éxito en la vacuna contra el VIH/Sida tras 25 años de pandemia. Esto demuestra que el enfoque de esperar y responder no es suficiente". Por este y otros motivos, GVF recoge sus datos de múltiples fuentes, incluyendo la investigación antropológica en los puntos calientes de enfermedades para identificar cómo los virus pasan de la vida silvestre a los seres humanos, y sigue las tendencias de las redes sociales para predecir y prevenir brotes. Utilizando una red de puestos de escucha viral localizados en Came-

rún, la República Democrática del Congo, Madagascar, China, Malasia, Sierra Leona y Gabón, GVF tiene como objetivo detener los virus antes de que pasen de animales a los humanos.

Para GVF, el mundo tanto físico como digital es el laboratorio, creando grandes conjuntos de datos que permiten tomar medidas preventivas reales. Pero siempre hay una necesidad de más datos para ayudar a predecir y prevenir la próxima pandemia mundial. GVF puede hacer predicciones precisas y realizables utilizando la ingeniería y la técnicas de software que fueron incipientes hace una década, pero para mejorar la eficacia de estas predicciones, esta institución está constantemente buscando nuevas fuentes de datos relevantes como crear modelos de interacción social a partir de los datos anónimos de teléfonos móviles.

Como resultado a su trabajo de campo, GVF ha averiguado cómo se propagan los virus de la carne de animales silvestres a los humanos; ha identificado una quinta forma de la malaria humana, y estudiado cómo se originó la enfermedad; y ha conseguido saber las causas del fallo en el control de la contención de la gripe porcina, entre otros éxitos.

Mejorar la eficiencia energética con la automatización

Silver Spring Networks ofrece a las compañías eléctricas un conjunto de hardware, software y servicios de smart grid. Sus productos se conectan a cada dispositivo en la red, creando una red inalámbrica entre dichos dispositivos que se puede personalizar a las necesidades e infraestructura de cada eléctrica. Utilizando una infraestructura de red inteligente, los dispositivos de Silver Spring crean una red segura de auto-configuración que garantiza la cobertura total a cualquier escala. En esta red, las eléctricas pueden implementar sistemas de medición avanzada para obtener datos procesables en programas de eficiencia energética, programas de respuesta a la demanda que optimizan la eficiencia energética mediante la reducción de la carga máxima, mejorando la eficiencia y la fiabilidad de sus servicios con la automatización de distribución.

A medida que ha ido aumentando la base de instalación de Silver Spring, ha ido creciendo su necesidad de contar con un data warehousing y una plataforma de análisis que le permita obtener una mejor visión y valor de sus datos. Por ello, ha cambiado a las soluciones de Enterprise Data Cloud de Greenplum para reemplazar su costoso hardware de back-end y aprovechar los beneficios de los análisis ágiles y profundos que la tecnología de Greenplum le proporciona. En concreto, Silver Spring emplea una combinación de productos SaaS y bajo demanda de Greenplum, para ofrecer servicios y productos de smart grid más fiables y eficientes. Además, este cambio ha permitido a Silver Spring generar nuevas fuentes de ingresos al poder ofrecer herramientas de análisis mejoradas para utilities que permiten conocer cómo los clientes utilizan la energía.



Un almacenamiento adecuado para emitir a más de 200 países

TV5MONDE es el canal internacional de televisión en lengua francesa que emite a un promedio diario de 25 millones de espectadores y 180 millones de hogares en más de 203 países. Con sede en París, siete canales diferentes de TV5MONDE ofrecen subtítulos en varios idiomas con el fin de llegar al público de forma simultánea en los cinco continentes. Tiene más de ocho centros de distribución primaria, emisiones a través de 52 satélites, más de 6.000 sistemas de cable, y disponibilidad en tres millones de habitaciones de hotel. TV5MONDE es el segundo canal distribuido más grande del mundo.

Para hacer frente a las miles de horas de programación original, control de versiones, y ampliar la catalogación de contenidos, TV5MONDE ha hecho una gran inversión en TI para gestionar la carga de trabajo de los grandes archivos

multimedia, gráficos y animación. Sin embargo, la creciente necesidad de emitir para formatos de pantalla ancha, transmisión de alta definición y televisión por IP, ha hecho que TV5MONDE necesitara aumentar bastante la capacidad y el rendimiento de su plataforma de almacenamiento. Para ello, en un principio pensó en actualizar la infraestructura SAN que tenía, pero las limitaciones en la escalabilidad y un rendimiento pobre llevaron al canal de televisión francés a explorar otras alternativas.

TV5MONDE ya había implantado el sistema de almacenamiento Isilon de EMC para utilizarlo en otra división. El éxito de esta instalación empujó al canal a llevar a cabo un piloto de almacenamiento NAS (Network-Attached Storage) con una solución mucho más grande de Isilon, que utilizó como nue-

va infraestructura de almacenamiento para la mayor parte de sus contenidos audiovisuales.

"Escogimos Isilon porque era fácil de implementar y es progresivo en términos de capacidad de almacenamiento y ancho de banda", comenta Guillaume Lairloop, director adjunto de Sistemas de TI para TV5MONDE. Asimismo, Isilon se integra a la perfección con el software SGT MAM permitiendo que se almacenen y cataloguen hasta 15.000 horas de vídeo digital.

EMC Isilon es la nueva plataforma de almacenamiento de TV5MONDE, el cual ahora puede ofrecer una mayor variedad de programación en alta definición. Además, este canal internacional de televisión tiene ahora la capacidad de poder aumentar terabytes en su infraestructura de almacenamiento en línea con la demanda.



www.tv5monde.com/latina

TV5MONDE América Latina y Caribe, la mejor programación y todo el entretenimiento desarrollado especialmente para el público latinoamericano. Ahora puedes disfrutar más horas de programación subtitulada, cubriendo el 50% de la grilla diaria, un nuevo sitio web con atractivos contenidos en todos los idiomas, el lanzamiento del pack VOD subtitulado con los programas más destacados y la creación de cursos de francés para debutantes. Una vez más, TV5MONDE pone a disposición de los televidentes lo mejor del entretenimiento y la cultura con propuestas para todos los gustos.

EL CANAL EN FRANCÉS PARA TI
TV5MONDE

Paloma Herranz, directora de Preventa para España de EMC

“Los datos son el nuevo centro de gravedad”

En la era del Big Data las empresas deben saber adaptarse con rapidez a los nuevos tiempos, y saber convertir todos los datos que genera la organización en información útil pa-

ra una correcta y ágil toma de decisiones. Para ello, Paloma Herranz, directora de Preventa de EMC en España, explica en esta entrevista cómo es necesario cambiar el enfoque

en cuanto a tecnologías, aplicaciones y profesionales para poder almacenar, gestionar y analizar todos los datos, los cuales han pasado a ser el centro de gravedad.

Big Data es un concepto relativamente nuevo que está en boca de toda la industria, ¿cómo explicaría este fenómeno desde el punto de vista de EMC?

Para nosotros es un enfoque nuevo de cómo almacenar, gestionar y analizar todos los datos que genera una organización para una óptima toma de decisiones. Estos datos son tanto estructurados como no estructurados y provienen de diferentes fuentes como el correo electrónico, imágenes médicas, dispositivos móviles, redes sociales... en definitiva, todas las nuevas fuentes que han surgido en los últimos años. Lo que entendemos por Big Data, y todas las tecnologías asociadas, es convertir los datos en información útil para la toma de decisiones en cualquier organización. Se

explotarlas. El análisis transaccional que se hace ahora de los datos es completamente diferente, así como las tecnologías que lo capturan, almacenan, gestionan y recuperan.

¿Cuál es entonces la oferta de EMC en el campo del Big Data?, ¿en qué áreas de producto se concreta?

Ofrecemos soluciones para toda la cadena de valor de los datos, es decir, desde la captura del dato hasta el almacenamiento, gestión, análisis de procesos y la consulta de la información. Todo ello con escalabilidad ya que podemos estar hablando ahora de un petabyte, y dentro de diez años, serán los zetabytes. Por ello, proporcionamos tecnología que permite, en todo su

Hay que conseguir un acercamiento proactivo al cliente mediante entornos analíticos fiables y escalables

trata de conseguir un acercamiento proactivo al cliente mediante entornos analíticos que sean fiables y escalables, y permitan crear nuevos productos, modelos de negocio... Y todo esto conlleva la adaptación de procesos y tecnologías porque hay que analizar los datos de una manera diferente.

¿Se puede gestionar el Big Data con las bases de datos relacionales y aplicaciones de Business Intelligence tradicionales?

Esa gestión se lleva haciendo desde hace muchos años, pero ahora se requiere adoptar un nuevo enfoque. Hay que cambiar los procesos de negocio y las bases de datos relacionales que antes teníamos, porque ahora los datos proceden de muchas fuentes diferentes, por lo que hay que utilizar tecnologías nuevas para

ámbito, gestionar diferentes modelos y escalabilidades en cuanto a los datos. Para EMC, “Data is the new gravity”, según nos explicaba Pat Gelsinger en EMC World en Las Vegas. Antes, las aplicaciones eran el centro de gravedad en todas las organizaciones, y ahora el modelo ha cambiado convirtiéndose los datos en el centro de gravedad. Bajo esta perspectiva, EMC apuesta, no solo por los conceptos de análisis de datos, también por cómo se capturan, se almacenan, o se mejoran los procesos internos de negocio, con productos que van desde soluciones como Greenplum para análisis de datos, a tecnologías como Isilon, Atmos, VNX... para almacenamiento; Data Domain y Networker para la protección de la información; RSA para seguridad; e IIG (la



familia de Documentum) para la mejora de la explotación de los procesos de negocio, y todo ello contando con tecnologías de virtualización.

¿Qué tipo de organismos o empresas pueden emplear tecnología de Big Data?

En principio, desde nuestro punto de vista, todos los sectores son susceptibles de utilizar la tecnología Big Data, desde el sector financiero, energético, telefonía móvil, Administración Pública, investigación... ¿quién no tiene datos para innovar y buscar nuevos modelos de negocio y más en los tiempos que corren! Por ejemplo,

nosotros tenemos proyectos en entidades financieras que, con tecnología de almacenamiento y análisis de información, tenían diferentes datos de clientes en distintas fuentes y sistemas transaccionales. Estos datos relativos se han agrupado de forma inteligente y esto ha permitido hacer una segmentación para poder ofrecer los productos más adecuados en función de la capacidad adquisitiva, su nivel de riesgo, inversiones a asumir... Igualmente ocurre en la Administración, que tiene multitud de datos y expedientes, que pueden entrecruzarse entre los distintos organismos, locales y centrales, para mejorar

el servicio al ciudadano. Hay estudios que desprenden que los gobiernos pueden ahorrar en Europa alrededor de 150 millones de dólares si gestionaran bien los Big Data.

Ha comentado que Big Data implica un cambio de enfoque en tecnología, porque se necesita una escalabilidad importante; y en aplicaciones pues el dato es el centro de gravedad. Pero, ¿cuál es el papel que juega el profesional en la evolución del Big Data?

Para las empresas el Big Data es todo un desafío a nivel tecnológico y de profesionales. Ha surgido un perfil nuevo, el

científico de datos, que realmente no existe; se tiene que formar para tender un puente entre la tecnología y el conocimiento del negocio porque los datos que se analizan tienen que dar un resultado y redundar en el beneficio del negocio. Por tanto, debe tener un conocimiento de los dos mundos. Es cierto que el científico de datos ha pasado a tener una relevancia especial y muchas organizaciones lo están resolviendo formando a gente que ha estado en tecnología y negocio. Es un reto organizativo muy importante que debe resolverse en el corto plazo.

¿Cómo ve EMC la evolución de esta tecnología en los próximos años?

El fenómeno Big Data acaba de empezar, la masa de datos está continuamente creciendo. Se dice que el universo digital crecerá un factor 44x en diez años. Esto ha sido posible porque se han unido diversas tecnologías como la virtualización, el cloud computing... conceptos nuevos que las empresas tienen que ser capaces de digerir. Nos estamos moviendo de contenido del dato al contexto. Las empresas tecnológicas deben ir en este camino y ser conscientes del fenómeno. Hay que invertir en tecnologías que ayuden a tener una mejor oferta y una mayor cohesión e integración de todos los elementos necesarios para rentabilizar el Big Data.

¿Qué conexión tiene el Big Data con el cloud computing?

Cloud es un elemento catalizador porque cuando se tiene la información en la nube es más sencillo almacenar y analizar el dato. El cloud computing es la base para poder ir a un modelo de Big Data del que se puede sacar mayor partido en el menor tiempo. El nivel de éxito en el Big Data también depende del nivel de madurez de las empresas, y la nube es un paso clave en la flexibilidad y la economía de coste.

Las herramientas sociales de Greenplum Chorus ya se encuentran disponibles

EMC casa el Big Data con las redes sociales empresariales para optimizar su analítica

Ya está disponible en el mercado Greenplum Chorus, una herramienta de colaboración social al estilo Facebook, que

EMC proporciona para que los profesionales responsables del análisis de datos la empleen en el estudio de conjuntos

de datos y faciliten de forma rápida y eficiente información útil para el negocio de sus organizaciones.

● Con su nueva plataforma de analítica para Big Data, Greenplum Chorus, EMC está acercando las capacidades de las redes sociales y herramientas de colaboración a los analistas de datos, llamados “científicos de datos”, un nuevo perfil recientemente creado para poder gestionar el fenómeno de la explosión de la información.

Greenplum Chorus es el primer conjunto de herramientas sociales de la industria para Big Data, que ha sido creado conjuntamente por la división Greenplum de EMC y los expertos de Pivotal Labs en modernas metodologías de desarrollo ágil y marcos de programación. Permite a los profesionales responsables de analizar la información colaborar en conjuntos de datos al estilo Facebook, con el objeto de permitir a las empresas adquirir una mayor visión y valor económico del Big Data.

“Chorus se sitúa en la intersección entre lo social y Big Data”, recalca Luis Zamora, responsable comercial de Greenplum en España y Portugal. “La unión de los conceptos social y colaboración ayuda a los grupos a la hora de colaborar y, de hecho, obtener aportaciones útiles a partir de sus datos. Se trata de un proceso muy social y colaborativo. Afecta a muchas personas dentro de una empresa. Mientras la mayor parte de la industria está centrada en la velocidad de carga, los tiempos de procesamiento, etc., EMC toma la iniciativa en la ‘socia-



lización’ del análisis de Big Data”, explica.

Una plataforma abierta

En la industria del software de hoy en día, contar con la implicación de los desarrolladores es un recurso que no tiene precio, y EMC lo sabe. Ha tomado la decisión de animar a los desarrolladores a que entren en su plataforma de analítica del Big Data al proporcionarles un entorno de desarrollo extremadamente ágil en el que trabajar y colaborar. Además, cierra sinergias con VMware al ofrecerles un entorno virtual independiente en el que pueden descargar un set de datos y trabajar con ellos sin interferir con otros analistas o incurrir en el peligro de corromper la información original. De esta manera, los científicos de datos pueden comentar, modificar o compartir los resultados con otros analistas a través de Chorus.

La gran oportunidad de analizar el Big Data

Hoy en día las empresas no tienen escasez de fuentes de datos, como lo demuestran los datos de IDC, que indican que en 2011 el mundo creó 1,8 zetabytes (1,8 billones de gigabytes) de información, lo que significa que la creación de datos se ha multiplicado por nueve en cinco años. Algunos de estos datos son estructurados, están almacenados en bases de datos y a menudo sirven como material para actividades de Business Intelligence. Debido a que el volumen de datos estructurados ha crecido, las bases de datos analíticas han madurado y continuado proporcionando una mejor analítica de datos estructurados. Sin embargo, el Big Data es un reto que va más allá de esta evolución porque la mayoría de los datos no son estructurados.

La conclusión es que aquellas empresas que buscan la agilidad del negocio necesitan una manera de analizar todos estos datos estructurados y no estructurados; y eso es un desafío que supera las capacidades de las bases de datos relacionales. Sin embargo, según un estudio de TDWI, el 70% de los profesionales de bases de datos consideran el Big Data una oportunidad, más que un problema. Sin embargo, pese a que cerca de tres cuartas partes de las organizaciones han adoptado alguna forma de analítica, sólo el 34% analiza el Big Data. Y las empresas que ya han adoptado este análisis, pueden hoy en día adelantarse a su competencia.

Lo más destacado de Greenplum Chorus

Colaboración en tiempo real: Chorus elimina el problema de los silos en una organización al reemplazar el correo electrónico acumulado por una interfaz única para todos los datos de la empresa, junto con bases de datos virtuales que favorecen la investigación e innovación, así como la colaboración. El resultado es que el equipo de científicos de datos puede, en colaboración, descubrir, compartir y discutir ideas que tienen un impacto significativo en el negocio.

Exploración de datos: Con Chorus, los analistas pueden navegar y explorar conjuntos de datos y bases de datos Hadoop de toda la organización. Una vez importados al entorno de trabajo, Chorus gestionará el flujo de datos y monitorizará la dependencia y actualización de la copia cuando se actualice la fuente con nueva información.

Espacios de trabajo: Con Chorus, los equipos de científicos de datos pueden crear nuevos entornos de analítica con unos cuantos clics. No es necesario hacer una solicitud al departamento de TI y esperar horas e incluso semanas.

Análisis de datos: Greenplum Chorus crea y mantiene un repositorio activo de soluciones analíticas, que facilita el intercambio de ficheros, el control de versiones, seguimiento de los cambios y archivado. Esta infraestructura de intercambio de contenido preserva el valor de los ficheros con independencia de los cambios en la organización.

Publicar puntos de vista: Chorus proporciona una única interfaz enriquecida con funcionalidades de redes sociales que facilitan la participación y colaboración entre los responsables de analizar el Big Data.

Greenplum Chorus es también una herramienta abierta, pues el data Warehouse de Greenplum está basado en la base de datos de código abierto PostgreSQL. Asimismo, se complementa con la iniciativa OpenChorus que EMC ha creado para acelerar la innovación y adopción de aplicaciones de datos colaborativas y sociales, ejecutadas bajo la plataforma Greenplum Chorus.

El código abierto de Greenplum Chorus se lanzará bajo licencia open source durante el segundo semestre de 2012; y las

actualizaciones del proyecto estarán disponibles en openchorus.org. No obstante, Greenplum Chorus también forma parte de la plataforma de EMC Greenplum Unified Analytics Platform (UAP), la base para que las empresas puedan crear su propia estrategia de analítica del Big Data. Esta plataforma combina tres productos de Greenplum en una oferta combinada: la base de datos para datos estructurados; la oferta empresarial Hadoop para analizar y procesar los datos no estructurados; y Chorus.

Redacción, publicidad, administración y suscripciones
Miguel Yuste 17-3ª 28037
Tel: 913 137 900
Fax: 913 273 704

Editorial

Director
Ambrosio Rodríguez • ambrosio.rodriguez@bps.com.es

Redactor jefe
Rufino Contreras • rufino.contreras@bps.com.es

Jefes de sección
Lores Serrano • lores.serrano@bps.com.es
Lucía Bonilla • lucia.bonilla@bps.com.es

Redactores y Colaboradores
Cristina López • cristina.lopez@bps.com.es
Ángela Mandianes • computing@bps.com.es
Humberto Minaya • redestelecom@bps.com.es

Diseño y maquetación
Jorge P. Aguilar • jorge.pariante@bps.com.es

Fotografía
Santiago Ojeda

Coordinador área de Internet
Fernando Reinlein • fernando.reinlein@bps.com.es

Marketing y Eventos
Responsable
Ruth González • ruth.gonzalez@bps.com.es

Producción
Imprime: PrinterMan | Cubiertas: AGS

Publicidad
Directora de Publicidad
Inma Pérez • inma.perez@bps.com.es
Tel.: +34 91 313 79 39 / Fax.: +34 91 327 37 04

COMPUTING España está editado por BPS

Consejera Delegada
Anunciación López • tita.lopez@bps.com.es

Depósito Legal M-7753-1995
ISSN 1139-9619



BIG DATA

TRANSFORMS BUSINESS



EMC²