



White Paper

IBM x3650 M4 BD: Servidor System x para análisis de Big Data

Por Kerry Dolan, Lab Analyst and Mark Bowker, Senior Analyst

Diciembre 2013

Este White Paper de ESG fue encargado por IBM y se distribuye con una licencia de ESG.

Índice

| | |
|---|---|
| La era de Big Data | 3 |
| Mejoras tecnológicas | 3 |
| ¿Qué pueden hacer por usted el análisis de Big Data? | 3 |
| Requisitos de los servidores de Big Data | 5 |
| Un servidor diseñado para las exigencias de Big Data: IBM x3650 M4 BD | 6 |
| La gran verdad | 8 |

Todas las marcas registradas son propiedad de sus respectivas empresas. La información que se brinda en esta publicación ha sido extraída de fuentes que The Enterprise Strategy Group (ESG) considera confiables, pero ESG no puede garantizarlo. Esta publicación puede contener opiniones de ESG, que están sujetas a modificaciones en cualquier momento. Esta publicación está protegida por derechos de autor cuyo titular es The Enterprise Strategy Group, Inc. Cualquier reproducción o redistribución de esta publicación, en parte o en su totalidad, ya sea en formato en papel, electrónico o de cualquier otro tipo, a personas que no están autorizadas para recibirla, sin el consentimiento expreso de The Enterprise Strategy Group, Inc., constituye una violación de la ley de derechos de autor de los EE.UU., y estará sometida a acciones civiles y, de ser aplicable, penales. Si tiene alguna pregunta, póngase en contacto con el departamento de relaciones con los clientes de ESG al teléfono 508.482.0188.

La era de Big Data

Todos los días se generan y se extraen datos: a través de transacciones online o en persona, búsquedas en Internet, artículos leídos, encuestas electrónicas, escaneos de códigos de barras, etc. Las empresas generan datos cuando rastrean inventarios, facturas, mediciones del desempeño y consumo de energía. Los clientes también proveen datos a las empresas, a través de terminales en los puntos de venta, del procesamiento de pagos, de la actividad en redes sociales y de grabaciones de audio y video. Incluso las máquinas generan datos, cuando los sensores y transmisores capturan información sobre el mantenimiento y el desempeño en miles de cosas, desde motores de aviones hasta turbinas de viento y trayectos de los camiones de entrega.

La cantidad de datos está explotando, tanto es así que lo llamamos Big Data. Estos conjuntos de datos traspasan los límites de las capacidades normales de procesamiento de datos. Para hablar de los volúmenes de datos ya no se usa solo el término terabytes, sino petabytes, exabytes, zettabytes e incluso yottabytes. Pero lo que hoy llamamos Big Data no sólo se refiere a grandes cantidades de datos, sino a datos en múltiples formatos, que no están en bases de datos estructuradas comunes. Estos datos semiestructurados y no estructurados (como web logs, archivos de video y audio, acción en las redes sociales, etc.) deben ser transformados de manera que puedan ser analizados fácil y rápidamente y, en última instancia, utilizados para impulsar los procesos de negocio.

Mejoras tecnológicas

Es únicamente por los avances tecnológicos que las empresas pueden realmente usar los datos por medio de plataformas de inteligencia empresarial y analíticas. Los avances en materia de potencia de procesamiento de los servidores, densidad de almacenamiento, velocidad de interconexión, arquitecturas de bases de datos y aplicaciones analíticas hacen posible obtener inteligencia útil a partir de esos datos y, a su vez, esta capacidad impulsa la extracción de más datos. Es como una bola de nieve que viene rodando por la ladera de una montaña: Cuando más rueda, más nieve junta, y cuanto más nieve tiene, más sigue rodando. Los volúmenes de datos y las mejoras tecnológicas se alimentan el uno del otro.

¿Qué pueden hacer por usted las aplicaciones analíticas de Big Data?

Lo que las plataformas analíticas como Hadoop, que es una plataforma de computación distribuida y de fuente abierta, construida para llevar a cabo analítica sobre grandes conjuntos de datos, le pueden brindar es un panorama mucho más amplio y completo de las cosas a las que se dedica su organización, ya sea clientes, reservas de gas y petróleo, estructuras celulares o métricas de desempeño. Por ejemplo, al analizar los archivos de registros de las interacciones bancarias online, las compañías de servicios financieros pueden comprender mejor qué es lo que sus clientes necesitan y cuándo lo necesitan. Los minoristas pueden comprender mejor las quejas de sus clientes analizando las grabaciones de los call-centers, y los análisis de las cámaras de seguridad pueden mejorar la seguridad de un campo. United Parcel Service, una de las empresas más grandes de envío de paquetes de todo el mundo, ha comenzado a reunir datos de sensores en sus vehículos, y los combina con datos en un mapa para analizar los recorridos de las entregas. ¿Qué gana con esto? Al reducir una milla por día y por conductor, la empresa cree que puede ahorrar unos 50 millones de dólares en combustible, mantenimiento de los vehículos y tiempo, además de reducir las emisiones de carbono.¹

En la investigación de ESG, más de la mitad (51 %) de los encuestados identificaron la reducción de costos como un beneficio clave de su plataforma analítica. Otros beneficios clave que citaron son: mejoras en los procesos, mejor agilidad del negocio, mayores ingresos, nuevas oportunidades, detección de fraudes, optimización del marketing y una mejor capacidad para controlar a los competidores.²

En última instancia, las empresas obtienen los siguientes beneficios:

- Ahorran tiempo y dinero.
- Responden con mayor rapidez a las oportunidades o a las amenazas del mercado.
- Interactúan con clientes en tiempo real.
- Mejoran la adquisición de clientes, su satisfacción y fidelidad.
- Identifican patrones y anomalías que mejoran la competitividad.

¹. Fuente: Artículo publicado en Bloomberg News, [UPS Crunches Data to Make Routes More Efficient, Save Gas](#), octubre de 2013.

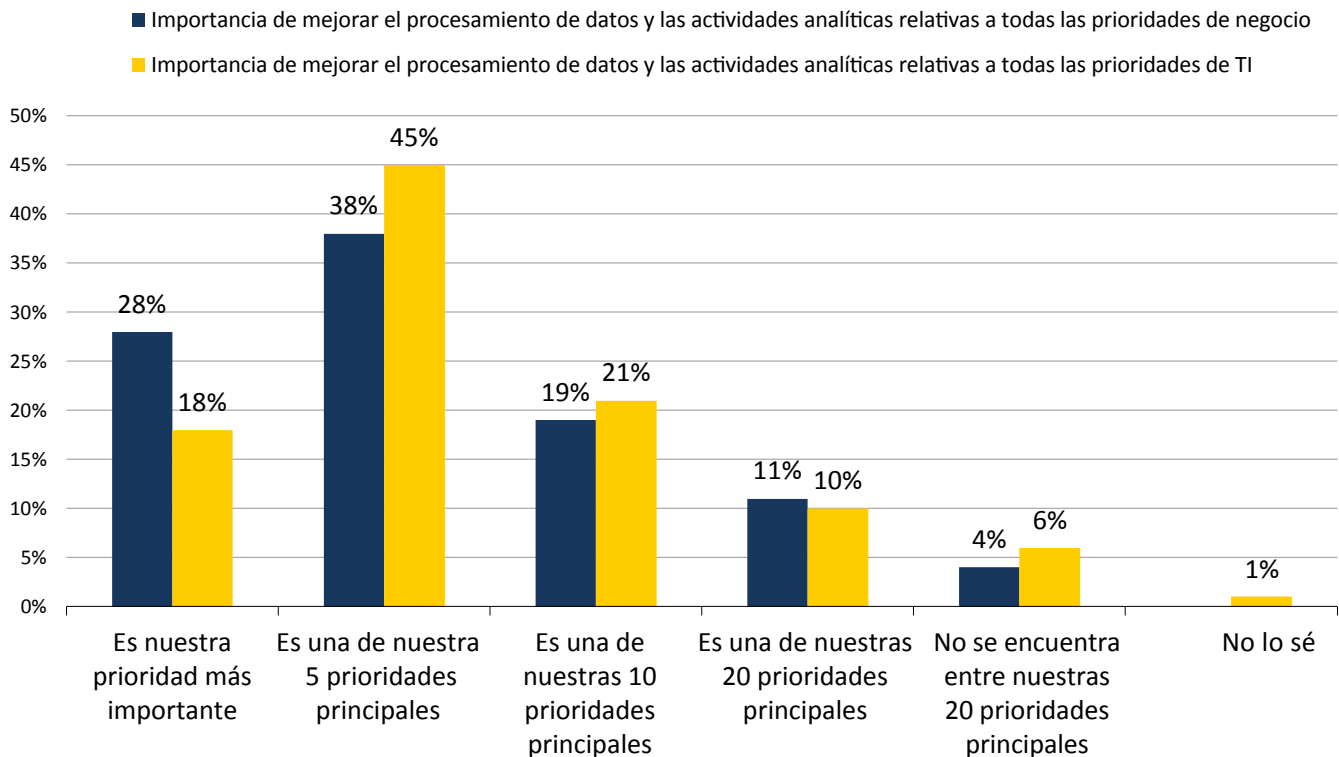
². Fuente: Reporte de investigación de ESG, [The Convergence of Big Data Processing and Integrated Infrastructure](#), julio de 2012. Todas las referencias y gráficos de la investigación de ESG que figuran en este documento han sido extraídos de este reporte, a menos que se especifique lo contrario.

Las analítica de datos cobra cada vez mayor importancia

Si tenemos en cuenta estos resultados potenciales, ¿cuál es la importancia de Business Intelligence y la analítica de datos? En la investigación de ESG con profesionales de negocio y de TI de EE.UU., el 28% de los encuestados calificaron la importancia de mejorar el procesamiento de datos y las actividades analíticas como la prioridad más importante con respecto a todas las prioridades de negocio, y otro 38% la calificó entre las cinco prioridades principales para la empresa (ver Figura 1). Esto significa que sólo la tercera parte de los encuestados lo consideran fuera de las cinco principales. La importancia que tiene el procesamiento y los análisis de datos para los departamentos de TI ha aumentado claramente en los últimos años: Según las encuestas anuales de ESG sobre intenciones de gastos en TI, y en términos de prioridades de TI, las iniciativas de Business Intelligence y de analítica de datos han cobrado mayor importancia en los últimos años.³

Figura 1: la importancia de mejorar las actividades de procesamiento de datos y analítica

Según todas las prioridades de su organización en materia de prioridades de negocio y de TI para los próximos 12-18 meses, ¿cómo calificaría la importancia de mejorar el procesamiento de datos y las actividades analíticas? (Porcentaje de encuestados, N=399)



Fuente: Enterprise Strategy Group, 2013

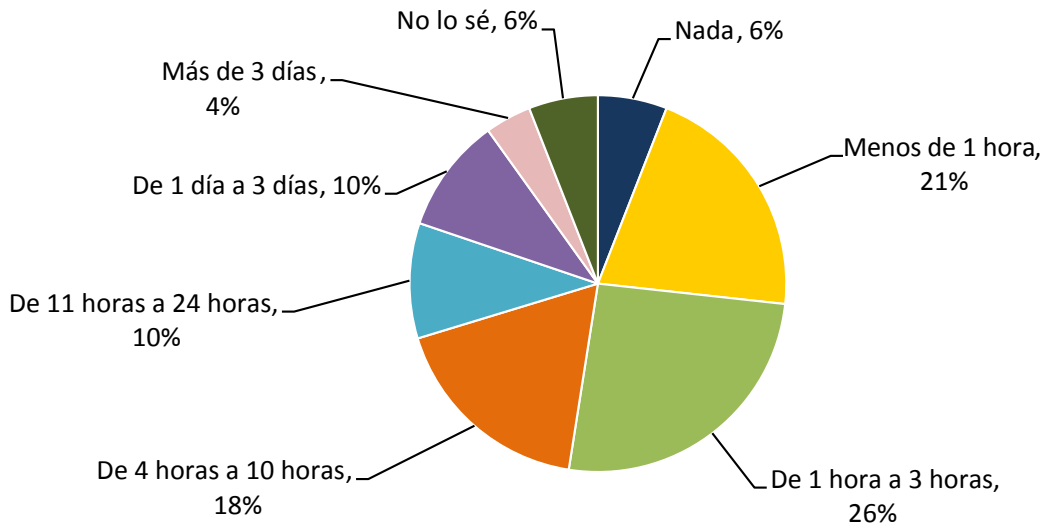
Tiempo de inactividad

A medida que el análisis de grandes volúmenes de datos cobra una importancia estratégica, cada vez más organizaciones valoran más su disponibilidad. Asimismo, con el aumento de la analítica en tiempo real, la disponibilidad resulta esencial. Como se puede apreciar en la Figura 2, el 53 % de los encuestados por ESG dijeron no ser capaces de tolerar más de tres horas de inactividad en la plataforma analítica sin una pérdida significativa de ingresos u otro impacto adverso al negocio, y el 6 % no puede tolerar ningún tiempo de inactividad.

³. Fuente: ESG Research Report, [2013 IT Spending Intentions Survey](#), enero de 2013.

Figura 2. Cantidad de tiempo de inactividad que pueden tolerar las plataformas analíticas de datos

Indique la cantidad de tiempo de inactividad que las plataformas de análisis de datos de su organización pueden tolerar antes de que su organización sufra importantes pérdidas de ingresos u otro impacto adverso para el negocio. (Porcentaje de encuestados, N=399)



Fuente: Enterprise Strategy Group, 2013.

Claramente, el análisis de grandes volúmenes de datos se está transformando rápidamente en una parte esencial de la estrategia corporativa. El retorno de su inversión en Big Data puede ser enorme si implementa la infraestructura y las aplicaciones más adecuadas, y si obtiene la ayuda que necesita. Por ende, la relación costo/desempeño para sus cargas de trabajo de Big Data es un parámetro importante a la hora de pensar en adquirir soluciones.

¿Está comprando la infraestructura y aplicaciones del tamaño correcto? ¿Existen arquitecturas de referencia que vengan preconfiguradas y validadas que puedan simplificar el proceso de compra e implementación? Estas preguntas deben tenerse en cuenta a la hora de evaluar soluciones analíticas de Big Data.

Requisitos de los servidores de Big Data

Según la investigación de ESG, más de la mitad (el 53 %) de las organizaciones encuestadas utilizan almacenamiento en servidores internos como un repositorio para análisis de datos. Pero hay un problema: Las técnicas tradicionales de procesamiento y análisis de datos con tecnologías de servidores convencionales no cuentan con la capacidad, el ancho de banda y la capacidad de procesamiento para manejar las exigencias de estos datos. Los procesos tradicionales fueron diseñados para datos estructurados en tamaños limitados, y no para diversos formatos de datos en grandes cantidades.

Para analizar estos conjuntos de datos, se necesitan dos funcionalidades clave: procesamiento de alta performance, y mucho almacenamiento. Algunos proveedores hablan de usar servidores “mercancía” para clústeres Hadoop. ¿Acaso esto significa que se puede comprar servidores baratos que no tienen los caballos de fuerza suficientes para impulsar actividades de procesamiento paralelo? Probablemente esta no sea una buena estrategia. ¿Eso significa que usted compra servidores que mezquinan el almacenamiento interno? Como el almacenamiento es esencial para distribuir tareas de procesamiento (en el caso de plataformas del tipo Hadoop), eso tampoco funciona.

La infraestructura de servidores debería satisfacer las necesidades de las tecnologías Hadoop y de Big Data. Esto suena bastante simple. Pero veamos el uso de la frase satisfacer las necesidades: a veces las organizaciones piensan que cuanto mayor es la infraestructura, es mejor para Big Data. En realidad, más no siempre significa mejor. Las infraestructuras de Big Data demasiado elaboradas son sólo un desperdicio de dinero. Como la analítica del tipo Hadoop depende de procesamiento distribuido utilizando clústeres de servidores, es importante que los servidores tengan la cantidad correcta de potencia, memoria, almacenamiento y características, pero no más de lo necesario.

ESG cree que el servidor que uno elige hace una gran diferencia. Puede comprar algo que no tenga demasiadas funcionalidades para hacer el trabajo bien y rápido. También puede usar algo pesado y demasiado elaborado. O bien, puede comprar el servidor de IBM que sea el correcto: Un servidor que incluya la cantidad correcta de capacidad de almacenamiento y desempeño, con espacio para crecer cuando su empresa esté lista. Un servidor que consuma poca energía y que esté altamente disponible.

Un servidor diseñado para las exigencias de Big Data: IBM x3650 M4 BD

IBM comprende las complejidades de las soluciones de Big Data y las aplicaciones analíticas, y ha diseñado un servidor específicamente para entornos de Big Data como Hadoop. El servidor System x3650 M4 BD (Big Data), que forma parte de la línea System x de IBM, es un servidor 2U en rack de dos sockets y construido especialmente para análisis de Big Data. Como muchas otras innovaciones de IBM, es más inteligente -como en Analítica inteligente, diseñada para la empresa más inteligente y un mundo más inteligente- con la combinación exacta de poder de procesamiento y capacidad de almacenamiento, además de ahorro de energía y alta disponibilidad. El x3650 M4 BD cuenta con una configuración de almacenamiento más inteligente y características que lo mantienen en pleno funcionamiento. Puede manejar cualquier trabajo de Big Data sin por eso dejar de mantener un perfil eficiente y económico.

Desempeño

En las implementaciones de Big Data, el desempeño es importante: los rápidos tiempos de respuesta permiten lograr mejores resultados de negocio y, por ende, las configuraciones de memoria y CPU hacen toda la diferencia. Así es como el IBM System x3650 M4 BD se ocupa del desempeño:

- **Poder de procesamiento.** Como estas aplicaciones se ejecutan en clústeres, usted necesita mucho CPU, pero no necesariamente el máximo poder de CPU en un único servidor. Sin embargo, la cantidad de núcleos es importante porque cuantos más haya usted puede configurar más ranuras MapReduce de Big Data por servidor. El x3650 M4 BD contiene potentes procesadores Intel Xeon E5-2600 v2, que ofrecen un desempeño inigualable, con hasta 12 núcleos por CPU.
- **Memoria.** Los servidores en un cluster Hadoop deberían aprovechar al máximo el ancho de banda de memoria. Este servidor incluye 16 ranuras DIMM y hasta 2DPC@1866 MHz, para hasta 512GB de memoria.

Almacenamiento

La capacidad de almacenamiento es otro ingrediente esencial de la analítica de Big Data en implementaciones Hadoop. Se necesita mucho almacenamiento porque Hadoop aporta desempeño y escalabilidad al distribuir las tareas de computación entre los discos en los nodos de los empleados. Además, cuantos más husos de disco haya, mejor podrá el sistema evitar acumulaciones de E/S en cualquier drive, lo que, a su vez, mejora el desempeño.

Muchos servidores que se utilizan en clústeres Hadoop tienen 12 drives JBOD para almacenar tanto el sistema operativo como los datos. En general, se necesitan dos drives para el sistema operativo. Eso deja al SO en drives no protegidos, lo que los hace vulnerables a las fallas, y a su vez son dos drives menos para el procesamiento de datos.

IBM ha diseñado este servidor de manera diferente. Incluye hasta 14 drives de 3,5 pulgadas. Usando drives de 4TB, lo que significa 56TB en un único servidor y, en el futuro, drives más grandes van a aumentar la capacidad potencial de almacenamiento, lo que reducirá aún más el bajo costo por terabyte. Es importante destacar que: Hay doce drives en el frente, por lo cual hay doce husos de disco disponibles para optimizar el desempeño. Asimismo, hay dos discos en la parte trasera para el sistema operativo (SO), por lo cual el SO no ocupa capacidad de almacenamiento que puede ser utilizada para procesar datos, y aquí se pueden usar drives más pequeños para minimizar los costos. Y lo más importante es que estos drives pueden tener protección RAID si se desea, para que el SO no sea vulnerable. Con el IBM x3650 M4 BD, usted obtiene la capacidad y desempeño más altos para los drives de datos, además de protección RAID y la reducción de costos para los drives del SO. (Como opción, se pueden configurar hasta 12Gbps de RAID para los drives de datos a fin de garantizar protección y para optimizar el RAID líder de la industria, y se pueden incluir hasta 4GB de caché Flash para mejorar el desempeño).

Energía

Un perfil de potencia más grande (y más costoso) de 130 vatios no es necesario para los servidores que están diseñados para clústeres. Los clústeres distribuyen el procesamiento, y se ejecutan en centros de datos con mecanismos de refrigeración incorporados. De todos modos, hay que equilibrar costo y energía. El x3650 M4 BD funciona a 95-115 vatios, tiene un amplio rango de temperatura operativa (que cumple con las especificaciones de eficiencia de ASHRAE), y alimentación eléctrica 80 PLUS Platinum de gran ahorro energético. Estas características permiten ahorrar mucha energía, en sintonía con la iniciativa “Un mundo más inteligente”, y ayudan a reducir el costo total de propiedad (TCO).

Alta disponibilidad

El análisis de Big Data tiene poca tolerancia al tiempo de inactividad, como se observa en la Figura 2, y si bien Hadoop es confiable en la gestión de datos, algunos agregados (como HBase para pares clave de valor) podrían dejar de funcionar si un nodo falla. Por eso, mantener estos servidores siempre funcionando resulta esencial. Además de una instalación y mantenimiento simples, IBM ha incorporado características avanzadas para la confiabilidad, disponibilidad y facilidad de servicio a este servidor, lo que promete un tiempo de actividad excepcional. Los drives y los suministros eléctricos redundantes e intercambiables sin necesidad de apagar el servidor son estándar, al igual que el diagnóstico de parches luminosos (que usan LEDs para encontrar la causa de los errores) y el Predictive Failure Analysis, que le permite detectar problemas antes de que interrumpan las operaciones. Además, la función IBM Feature on Demand le permite activar nuevas características de servicio de manera remota usando claves de usuario, sin interrupciones de funcionamiento. Esto es diferente de los agregados de características para servidores estándar, en los que hay que apagar el servidor y abrir la caja. Por último, la configuración de puertos es flexible, con soporte para Ethernet de 1Gb y de 10Gb. También cuenta con un puerto dedicado IMM2 de 1GbE, que integra el procesador de servicio, el controlador de video, la gestión remota y otras funcionalidades dentro de la serie System x.

Beneficios adicionales de IBM

Además de construir este servidor para trabajar con Big Data, IBM ofrece otros beneficios para cualquier solución de Big Data. Por ejemplo, el software InfoSphere BigInsights de IBM facilita el uso de Hadoop y la construcción de aplicaciones de Big Data. Agrega características administrativas, de descubrimiento, desarrollo, aprovisionamiento y seguridad, además de funcionalidades analíticas de IBM Research. Y el IBM Big Data System Center contiene más de 100 expertos en la materia para brindar servicios tales como el dimensionamiento de soluciones y configuraciones, la optimización de infraestructuras, el diseño de arquitecturas, la transferencia de conocimientos, etc. En resumen, sus funcionalidades de gran capacidad, alto desempeño y bajo costo por terabyte convierten el x3650 M4 BD en una poderosa solución para Big Data, con las características que usted necesita, y no las que no necesita. Cabe destacar que otras aplicaciones también se pueden beneficiar con estas características, tales como video on-demand, gestión de datos e implementaciones en la nube o de virtualización.

La gran verdad

Los avances tecnológicos han hecho posibles el acceso y el análisis de que antes no se podían obtener. Desde cotejar genomas hasta analizar video y enviar ofertas personalizadas a teléfonos móviles individuales, todo es posible gracias a las aplicaciones analíticas para Big Data. Las empresas de todos los sectores utilizan la analítica para tomar mejores decisiones de negocio, optimizar procesos, mejorar la rentabilidad, investigar fenómenos científicos y comprender mejor el mundo en que vivimos. Si desea ver ejemplos, sólo tiene que visitar el sitio Web de IBM, donde encontrará más de 100 estudios de casos acerca de sus clientes de analítica.

Como resultado, la analítica de datos está escalando posiciones en la lista de prioridades empresariales. El retorno de la inversión en su analítica de Big Data se basa en los recursos que tiene, y especialmente en el caso de Big Data, su infraestructura de servidores es un mecanismo fundamental dentro del proceso. Los ejecutivos de negocios exigen más de sus datos que nunca antes, y desean obtener respuestas más rápidas y de mayor impacto. Es esencial contar con la infraestructura adecuada.

El servidor x3650 M4 BD de IBM fue diseñado teniendo en cuenta las necesidades de las aplicaciones de Big Data. Ofrece alto desempeño, procesamiento de alta capacidad que resulta eficaz para datos estructurados y no estructurados, y mantiene a los servidores en funcionamiento con características de alta disponibilidad. La analítica en tiempo real, las implementaciones de Hadoop y las aplicaciones para extraer, transformar y cargar, todas se benefician con estas funciones.

El portafolio de System x de IBM incluye soluciones de servidores adaptadas a cargas de trabajo de nivel inicial, del negocio en general, críticas para el negocio y de alta densidad, y el nuevo x3650 M4 BD contribuye a la reputación de IBM para construir sistemas que cumplen con los requisitos específicos de las cargas de trabajo. Al igual que con otras áreas de necesidad (como computación en la nube, gestión de datos, HPC y escritorios virtuales), IBM ofrece hardware, software, arquitecturas de referencia y servicios para aplicaciones analíticas de Big Data e implementaciones Hadoop. IBM es una empresa reconocida por la confiabilidad de los sistemas, la satisfacción de los clientes y el soporte de proveedores. Por eso, si usted va a utilizar analítica de Big Data para ayudar a tomar decisiones de negocio, puede confiar en IBM y en sus socios para brindarle los sistemas y soluciones más adecuados.

IBM hace que más personas, productos, servicios, alianzas se acerquen y tomen conciencia de la importancia de las aplicaciones de Big Data. No estamos exagerando si decimos que la industria de Big Data le debe mucho a IBM. Por eso mismo, ESG cree que si usted está buscando un servidor para Big Data (u otras soluciones y servicios relacionados), IBM podría ser su respuesta, desde el principio hasta el final.



Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel: 508.482.0188 Fax: 508.482.0218 | www.esg-global.com