



Analíticas oscuras Iluminando las oportunidades ocultas en los datos no-estructurados

Esta es una traducción al español de la versión oficial en inglés del **Tech Trends 2017 – The kinetic Enterprise** (<https://www2.deloitte.com/global/en/pages/technology/articles/tech-trends.html>) – **Dark analytics. Illuminating opportunities hidden within unstructured data** - Pg. 20-33.
Traducción realizada por Samuel A. Mantilla, asesor de investigación contable de Deloitte & Touche Ltda., Colombia, con la revisión técnica de César Cheng, Socio Director General de Deloitte & Touche Ltda., Colombia

Analíticas oscuras

Iluminando las oportunidades ocultas en los datos no-estructurados

A través de las empresas, los almacenamientos de datos cada vez en más en ampliación permanecen no-estructurados y no-analizados. Pocas organizaciones han sido capaces de explorar las fuentes no-tradicionales de datos tales como archivos de imágenes, audio, y videos; el torrente de información de máquina y sensores generado por el Internet de las Cosas; y los enormes depósitos de datos brutos que se encuentran en lo más hondo de la “red inexplorada.” Sin embargo, los avances recientes en visión de computador, reconocimiento de patrones, y analíticas cognitivas están haciendo posible que las compañías arrojen una luz en esas fuentes no explotadas y deriven conocimientos que lleven a mejores experiencias y a mejor toma de decisiones a través de los negocios.

En esta era de ilustración orientada-a-la-tecnología, los datos son nuestra moneda competitiva. Sepultados por información bruta generada en volúmenes alucinantes por sistemas transaccionales, medios de comunicación social, motores de búsqueda, y otras tecnologías incontables están críticos conocimientos estratégicos, del cliente, y operacionales que, una vez iluminados por las analíticas, pueden validar o aclarar supuestos, informar la toma de decisiones, y ayudar a definir senderos nuevos para el futuro.

Hasta recientemente, tener un enfoque pasivo, que mira atrás, para los datos y la analítica, era la práctica estándar. Con la meta última de “generar un reporte,” las organizaciones frecuentemente aplicaron las capacidades analíticas a muestras limitadas de datos estructurados aislados en un sistema específico o en una función de la compañía. Por consiguiente, problemas molestos de calidad con datos maestros, carencia de sofisticación del usuario, e incapacidad para reunir datos provenientes de los sistemas de la empresa a menudo conspiraban para producir conocimientos que fueran, en el mejor de los casos, de alcance limitado y en el peor, conducentes a engaño.

En el presente, los CIO aprovechan arquitectura distribuida de datos, procesamiento en-la-memoria, aprendizaje de máquina, visualización, procesamiento natural del lenguaje, y analíticas cognitivas para resolver preguntas e identificar patrones y conocimientos valiosos que habrían parecido inimaginables hace solo unos pocos años. Además, las analíticas ahora dominan las agendas

y la velocidad de la TI. En la *2016 Global CIO Survey*, de Deloitte, realizada a 1,200 ejecutivos de TI, quienes respondieron identificaron a las analíticas como la principal prioridad para la inversión. De igual manera, identificaron contratar talento de TI con habilidades analíticas como su principal prioridad para el reclutamiento en los próximos dos años.¹

Aprovechando esos conjuntos avanzados de herramientas y habilidades, en los próximos 18 a 24 meses un creciente número de CIO, líderes de negocios, y científicos de datos estarán experimentando con las “analíticas oscuras”: exploraciones centradas en el enorme universo de datos no-estructurados y “oscuros,” con la meta de desenterrar el tipo de conocimientos altamente matizados sobre negocios, clientes, y operaciones que los activos de datos estructurados actualmente en su posesión no pueden revelar.

En el contexto de los datos de negocio, “oscuro” describe algo que está oculto o no-digerido. Las analíticas oscuras se centran principalmente en datos brutos basados-en-texto que todavía no han sido analizados – con un énfasis puesto en los datos no-estructurados, lo cual puede incluir cosas tales como mensajes de texto, documentos, correos electrónicos, archivos de video y audio, e incluso imágenes. En algunos casos, las exploraciones de las analíticas oscuras también podrían tener como objetivo la red profunda, la cual comprende todas las cosas que estén en línea y que no estén indexadas por motores de búsqueda, incluyendo el sub-conjunto pequeño de sitios anónimos, inaccesibles, conocidos como la “red oscura.” Es imposible calcular de manera exacta el tamaño de la

red profunda, pero algunos estiman que es 500 veces más grande que la superficie de la red que la mayoría de las personas buscan diariamente.²

En un clima de negocios donde los datos son moneda competitiva, esos tres recursos ampliamente inexplorados probarían ser similares a un premio de lotería. Aún más, los datos y los conocimientos allí contenidos se están multiplicando a una ritmo alucinante. Se estima que el 90 por ciento de los datos hoy disponibles fue generado durante los últimos cinco años.³ El universo digital – que comprende los datos que creamos y copiamos anualmente – está doblando su tamaño cada 12 meses. Además, se espera que llegue a 44 zettabytes (esto es, 44 *trillones* de gigabytes) en tamaño para el 2020 y contendrá casi tantos bits digitales como hay estrellas en el universo.⁴

Aún más, esas proyecciones actualmente pueden ser conservadores. Gartner Inc. Anticipa que el crecimiento explosivo del Internet de las Cosas [Internet of Things (IoT)] verá conectados a 20.9 billones de dispositivos desplegados para el 2020.⁵ En la medida en que el IoT se expanda, lo harán los volúmenes de datos que la

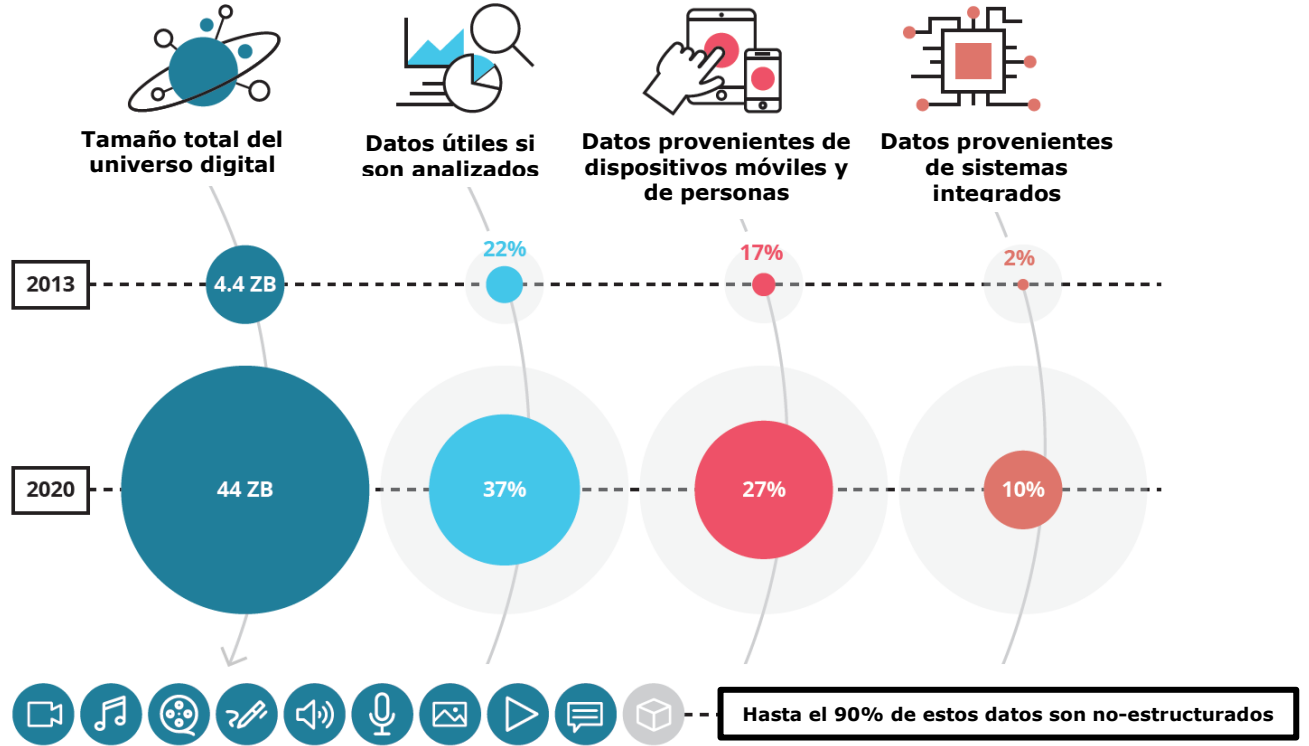
tecnología genera. Para algunos estimados, los datos que los dispositivos de IoT crearan globalmente en el 2019 – gran parte de los cuales serán “oscuros” – será 269 veces mayor que la cantidad de datos que están siendo transmitidos a los centros de datos desde los dispositivos de usuario final y 49 veces mayor que el total del tráfico de los centros de datos.⁶ Contra este trasfondo estadístico, los grandes datos, como un imperativo del negocio, pueden ser más exactamente descritos como “datos enormes.”

Hasta la fecha, las compañías han explorado solo una minúscula parte del universo digital para el valor analítico. El IDC estima que para el 2020, tanto como el 37 por ciento del universo digital contendrá información que puede ser valiosa si es analizada.⁷

¿Pero exactamente qué tan valiosa? El IDC también proyecta que las organizaciones que analicen todos los datos relevantes y entreguen información útil, para el 2020 lograrán \$430 billones extra en ganancias de productividad sobre sus pares menos orientados analíticamente.⁸

Figura 1. El universo digital en expansión, 2013 – 2020

En el año 2020, el universo digital se espera que llegue a 44 zettabytes. Un zettabyte es igual a un billón de terabytes. Los datos valiosos para las empresas, especialmente los datos no-estructurados provenientes del Internet de las Cosas y de fuentes no-tradicionales, está proyectado se incrementen en tamaños absolutos y relativos.



Fuentes: EMC Digital Universe with research and analysis by IDC, “The digital universe of opportunities: Rich data and the increasing value of the Internet of Things,” April 2014; International Data Corporation, “IDC iView: Extracting value from chaos,” 2011, www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf, accessed December 29, 2016.

Que haya luz

Cuando pensamos acerca del potencial de las analíticas, las posibilidades que visualizamos están limitadas a los datos estructurados que existen dentro de los sistemas que nos rodean. Las analíticas oscuras buscan remover esos límites mediante fundir una red mucho más amplia de datos que podamos capturar como un cuerpo de señales actualmente sin explorar.

Los esfuerzos de las analíticas de datos típicamente se centran en tres dimensiones:

Datos sin explotar ya en nuestra posesión: En muchas organizaciones, todavía están ociosas grandes colecciones de datos estructurados y no-estructurados. Por el lado estructurado, ello es típicamente a causa de que no ha sido fácil hacer las conexiones entre los conjuntos dispares de datos que puedan tener significado – especialmente información que vive fuera de un sistema, función, o unidad de negocios dado. Por ejemplo, una compañía de seguros grande mapeó las direcciones de las viviendas de los empleados y las asignaciones de parqueadero con sus calificaciones de la satisfacción del lugar de trabajo y la retención de datos. El esfuerzo reveló que uno de los mayores factores que incentivaban la rotación voluntaria era el tiempo de viaje – la combinación de distancia a la oficina, patrones de tráfico basados en la programación de cambio de los trabajadores, grado de dificultad en encontrar parqueadero, y la extensión del camino desde el carro hasta su lugar de trabajo.

En relación los datos no-estructurados “tradicionales,” piense en correos electrónicos, notas, mensajes, documentos, inicios de sesión, y notificaciones (incluyendo las provenientes de dispositivos de IoT). Esos están basados en texto y se encuentran dentro de las fronteras organizacionales pero permanecen ampliamente sin ser aprovechados, ya sea a causa de que no viven en una base de datos relacional o porque hasta relativamente reciente, no existían las herramientas y técnicas que se necesitan para aprovecharlas de manera eficiente. Sepultada en esos activos de datos no-estructurados podría estar información valiosa sobre fijación del precio, comportamiento del cliente, y competidores. Particularmente en compañías multinacionales, también pueden contener activos de datos no-traducidos potencialmente valiosos creados para o generados en mercados que no hablan inglés.

¿Qué porcentaje de los datos existentes hoy son no-estructurados? No lo sabemos con seguridad. La cifra generalmente aceptada ha sido el 80 por ciento – conocida como “la regla del 80 por ciento” – si bien estimados recientes colocan el número cerca al 90 por ciento.⁹

Datos no-estructurados no-tradicionales: La segunda dimensión de las analíticas oscuras se centran en una categoría diferente de datos no-estructurados a la cual no se le puede hacer minería usando las técnicas tradicionales de presentación de reportes y de analíticas – archivos de audio y video e incluso imágenes, entre otras. Usando visión de computador, reconocimiento avanzado de patrones, y analíticas de video y sonido, las compañías pueden ahora hacer minería de los datos contenidos en formatos no-tradicionales para entender mejor a sus clientes, empleados, operaciones, y mercados. Por ejemplo, un minorista puede ganar un entendimiento más matizado del humor o la intención del cliente mediante analizar imágenes de video de la postura, las expresiones faciales, o los gestos de los clientes. Una compañía de petróleo y gas podría usar sensores acústicos para monitorear los gasoductos y algoritmos para proporcionar visibilidad en las tasas y en la composición del flujo del petróleo. Un parque de diversión podría ganar mayor conocimiento sobre las demografías de los clientes mediante analizar la huella de las cámaras de seguridad para determinar qué tantos clientes llegan en carro, en transporte público, o a pie, y en qué momentos durante el día.

Cámaras de alta fidelidad envolvente de bajo costo, áreas de micrófonos de alcance largo, y cámaras de alta definición hacen posible monitorear todas las actividades de negocio que ocurren en una empresa. La capacidad para aplicar las analíticas a alimentaciones de audio y video en tiempo real abra profundas nuevas oportunidades para la detección de y la respuesta a señales. Tales señales digitales ofrecen nuevas maneras para responder preguntas existentes y para explorar nuevas oportunidades. Por otra parte, en los últimos años los costos de almacenamiento de datos han declinado en un estimado del 15 a 20 por ciento, haciendo del archivo de imágenes y registros de audio una opción más realista para las organizaciones más pequeñas.¹⁰

Datos en la red profunda: Como una dimensión de las analíticas oscuras, la red profunda ofrece lo que para muchos puede contener el cuerpo más grande de información sin explorar – datos curados por académicos, consorcios, agencias del gobierno, comunidades, y otros dominios de terceros. Pero el tamaño completo del dominio y la distinta carencia de estructura puede hacer difícil la búsqueda. Por ahora, solo los esfuerzos de minería de datos y de analíticas que estén delimitados y centrados en un objetivo específico – por ejemplo, datos con licencia de propiedad de una asociación privada – probablemente ofrecerán conocimientos útiles, relevantes. Así como la comunidad de la inteligencia monitorea el volumen y el contexto de la actividad de la red profunda para identificar amenazas potenciales, los negocios pronto pueden ser capaces de curar la inteligencia competitiva mediante usar una variedad de herramientas emergentes de búsqueda diseñadas para ayudar a los usuarios a definir investigación científica objetivo, datos de activistas, o incluso hilos de aficionados que se encuentren en la red profunda. Por ejemplo, Deep Web Technologies elabora herramientas de búsqueda para recuperar y analizar datos que estarían inaccesibles para los motores estándar de búsqueda.¹¹ Su software actualmente está siendo desplegado por agencias federales científicas así como también por varias organizaciones académicas y corporativas. Stanford University ha elaborado un prototipo de motor denominado Hidden Web Exposer que araña la red profunda por información usando un

enfoque de tarea-específica, asistido-por-humano. Otros motores de búsqueda públicamente accesibles incluyen Infoplease, PubMed, e Infomine de la University of California.¹²

Linterna, no estrella interplanetaria

Que quede claro, el propósito de las analíticas oscuras no es catalogar volúmenes enormes de datos no-estructurados. Fundir datos amplios de la red sin tener en mente un propósito específico probablemente llevará al fracaso. Además, los esfuerzos de las analíticas oscuras que son quirúrgicamente precisos tanto en intención como en alcance a menudo entregan el mayor valor. Al igual que cada camino de las analíticas, los esfuerzos exitosos comienzan con una serie de preguntas específicas. ¿Qué problema va usted a resolver? ¿Qué haríamos de manera diferente si pudiéramos resolver ese problema? Finalmente, ¿qué fuentes de datos y qué capacidades analíticas nos ayudarán a responder las primeras dos preguntas?

Responder esas preguntas hace posible que las iniciativas de analíticas oscuras iluminen conocimientos específicos que sean relevantes y valiosos. Recuerde, la mayoría del universo de datos es oscura, y con su tamaño y variedad complejos, probablemente continuará de esa manera.



Rx de IU Health para la minería de datos oscuros

Como parte del nuevo modelo de atención en salud, Indiana University Health (IU Health) está explorando maneras para usar datos no-tradicionales y no-estructurados para personalizar la atención en salud para los pacientes individuales y mejorar los resultados generales de la salud para la población más amplia.

La relación tradicional entre los proveedores de atención en salud y los pacientes a menudo son de naturaleza transaccional, centrándose en visitas individuales y resultados específicos más que prestar servicios holísticos de atención sobre una base continua. IU Health ha determinado que la incorporación de los conocimientos provenientes de datos adicionales ayudará a construir lealtad del paciente y prestará atención más útil, sin problemas, y costo-eficiente.

“IU Health necesita un entendimiento de 360 grados de los pacientes que sirva con el fin de crear un tipo de atención y servicios que les ayudará a mantenerse en el sistema,” dice Richard Chadderton, vicepresidente senior, compromiso y estrategia, IU Health. “Nuestra organización está explorando maneras para hacer minería de y analizar datos – de una manera muy similar a como las compañías orientadas-al-consumidor están enfocando los datos de los clientes – para desarrollar este entendimiento profundo.”¹³

Por ejemplo, considere las voluminosas notas libres de formatos – tanto escritas como verbales – que los médicos generan durante las consultas del paciente. Desplegar reconocimiento de voz, aprendizaje profundo, y capacidades de análisis de textos a esas fuentes a mano pero previamente sub-utilizadas potencialmente podría agregar más profundidad y detalle a los registros médicos del paciente. Esas mismas capacidades también pueden ser usadas para analizar grabaciones de audio de conversaciones del paciente con los centros de llamadas

de IU Health para mejorar adicionalmente los registros del paciente. Tales conocimientos podrían ayudar a IU Health a desarrollar un entendimiento más profundo de las necesidades del paciente, y a iluminar de mejor manera cómo esos pacientes utilizan los servicios del sistema de salud.

Otra oportunidad involucra usar datos oscuros para ayudar a predecir la necesidad a administrar la atención a través de las poblaciones. IU Health está examinando cómo la computación cognitiva, los datos externos, y los datos del paciente podrían ayudar a identificar patrones de enfermedad, acceso a la atención en salud, y resultados históricos en las poblaciones locales. Los enfoques podrían hacer posible incorporar factores socio-económicos que puedan afectar el compromiso de los pacientes con los proveedores de atención en salud.

“Puede haber una correlación entre alta densidad por unidad de vivienda y desconexión de la salud,” dice Mark Lantzy, vicepresidente senior y director de información jefe, IU Health. “Es prometedor que podamos aumentar los datos del paciente con datos externos para determinar cómo comprometernos mejor con las personas acerca de su salud. Nosotros estamos creando la plataforma subyacente para descubrir esas correlaciones y estamos intentando crear algo más sistémico.

El destino de nuestro camino es una experiencia mejorada del cliente,” continúa. “En últimas, lo que queremos es manejar mejores satisfacción y compromiso.

Más que entregar gran atención de salud a los pacientes individuales, también queremos mejorar la salud de la población de toda Indiana. Para poder impactar de alguna manera, incluso incrementalmente, seríamos enormemente benéficos.”¹⁴

Los minoristas lo hacen personal

Casi de manera universal los minoristas reconocen que lo digital le ha dado nueva forma al comportamiento del cliente y a las compras. De hecho, \$0.56 de cada dólar gastado en un almacén está influenciado por una interacción digital.¹⁵ Aun así muchos minoristas – particularmente los que tienen operaciones de ladrillo-y-cemento – todavía se esfuerzan por entregar las experiencias digitales que los clientes esperan. Algunos se centran excesivamente en sus competidores en lugar de en sus clientes y confían en los mismos viejos indicadores clave de desempeño y en los mismos datos viejos.¹⁶

Sin embargo, en los últimos años números crecientes de minoristas han comenzado a explorar diferentes enfoques para el desarrollo de experiencias digitales. Algunos están analizando datos previamente oscuros de las vidas de los clientes y están usando los conocimientos resultantes para desarrollar estrategias de mercadeo, comercialización, servicio al cliente, e incluso desarrollo de producto que les ofrecen a los compradores una experiencia específica e individualizada para el cliente.

Stitch Fix, es un servicio en línea de compra por suscripción que usa imágenes provenientes de medios de comunicación social y otras fuentes para hacerle seguimiento a las tendencias emergentes de la moda y evaluar las preferencias del cliente. Su proceso comienza con clientes que responden un cuestionario detallado acerca de sus gustos relacionados con la ropa. Luego, con permiso del cliente, el equipo de 60 científicos de datos

de la compañía aumenta esa información mediante escanear imágenes de los clientes en Pinterest y otros sitios de medios de comunicación social, los analiza, y usa los conocimientos resultantes para el desarrollo de un entendimiento más profundo del sentido de estilo de cada cliente. Los estilistas de la compañía y los algoritmos de inteligencia artificial usan esos perfiles para seleccionar los elementos de vestuario apropiados-para-el-estilo para ser enviados a clientes individuales con intervalos regulares.¹⁷

Mientras tanto, la cadena de supermercados de comestibles Kroger Co. está tomando un enfoque diferente que aprovecha el Internet de las Cosas y las técnicas analíticas avanzadas. Como parte de un programa piloto, la compañía está insertando una red de sensores y analíticas en las estanterías de los almacenes, los cuales pueden interactuar con la aplicación de Kroger y una lista de compras digitales en el teléfono del cliente. Cuando el cliente va por cada pasillo, el sistema – que contiene una historia digital de las compras del cliente y de las preferencias del producto – puede destacar, en pantallas de 4 pulgadas colocadas en los pasillos, productos con precio especial que el cliente puede querer. Este piloto, que comenzó a finales del 2016 con prueba inicial en 14 almacenes, se espera que se expanda en el 2017.¹⁸

En los próximos meses se espera ver más pilotos y despliegues completos como esos, en la medida en que los minoristas comiencen a ejecutar estrategias de compromiso del cliente que, si son exitosas, podrían transformar tanto la experiencia de compra como el rol que los datos no-tradicionales juegan en la industria minorista.

MI PARTE

GREG POWERS, VICE PRESIDENT DE TECHNOLOGY

HALLIBURTON

Como líder en la industria de servicios para campos petroleros, Halliburton tiene una larga historia de confiar fuertemente en los datos para entender las condiciones de operación actuales y pasadas en el campo y para medir el rendimiento en el pozo.

Sin embargo, el enorme volumen de información que podemos obtener y que obtenemos va mucho más allá del ancho de banda cognitivo humano. Los avances en la ciencia de sensores están entregando conjuntos enormes tanto de datos oscuros como de lo que pienso realmente son datos oscuros. Por ejemplo, nosotros podemos escanear electromagnéticamente las rocas para determinar su consistencia. Nosotros usamos resonancia magnética nuclear para realizar lo que equivale a una imagen de resonancia magnética en los pozos de petróleo. El análisis de neutrones y rayos gama mide la admisibilidad eléctrica y la conductividad de la roca. La espectroscopia del fondo del pozo mide los fluidos. Sensores acústicos recaudan diariamente 1-2 terabytes de datos. Todos estos datos oscuros nos ayudan a entender mejor el desempeño del pozo. De hecho, hay mucho valor potencial sepultado en esta oscuridad al cual yo le quito el marco y me refiero como "datos brillantes" que todavía tenemos para aprovechar.

Nosotros hemos hecho un buen trabajo de elaborar una vista retrospectiva del desempeño pasado. En la siguiente fase del programa continuo de analíticas de Halliburton, queremos desarrollar la capacidad para capturar, hacer minería, y usar los conocimientos de los datos brillantes para que se vuelvan más predictivos. Dada la naturaleza de nuestras operaciones, esto no será tarea pequeña. Eventos idénticos orientados por circunstancias comunes son raros en la industria de petróleo y gas. Nosotros tenemos 30 años de datos retrospectivos, pero hay un número infinito de combinaciones de roca, gas, petróleo, y otras variables que afectan nuestros resultados. Desafortunadamente, no hay una ecuación física constituyente de amplio alcance con la cual tratar cada situación que se encuentre. Aun así, si no podemos explicar lo que históricamente hemos visto, podemos explorar qué ha ocurrido y nuestra apreciación refinada

de los datos históricos sirve como una hoja de ruta hacia dónde podemos ir. En otras palabras, nosotros planeamos correlacionar los datos con cosas que estadísticamente parece que importan y, luego, usar estos datos para desarrollar un umbral de confianza para informar cómo debemos enfocar esos problemas.

“EN ESTA OSCURIDAD HAY MUCHO VALOR POTENCIAL SEPULTADO.”

Nosotros consideramos que los datos no-tradicionales tienen la clave para crear capacidades avanzadas de respuesta inteligente para resolver problemas, potencialmente sin intervención humana, antes que ocurran. Sin embargo, la industria de petróleo y gas de manera justificable es conservadora cuando se trata de adoptar nueva tecnología, y cuando se trata de automatizar nuestro manejo de la infraestructura crítica, la industria será más conservadora que como es usual. Esa es la razón por la cual vemos que surge un enfoque por niveles para aprovechar nuevas líneas de producto, herramientas, y ofertas. En el nivel más bajo, haremos mediciones y contaremos algo luego que el hecho haya ocurrido. En el siguiente nivel, nuestra meta será reconocer que algo ha ocurrido y, entonces, entender por qué ocurrió. En el siguiente paso usaremos monitoreo en tiempo real para proporcionar conciencia en-el-momento de lo que esté ocurriendo y por qué. En el siguiente nivel, las herramientas predictivas nos ayudarán a discernir qué probablemente ocurrirá luego. La oferta más extrema involucrará automatizar la respuesta - removiendo por completo de la ecuación la intervención humana.

La perforación es un trabajo complicado. Para hacerlo más autónomo y eficiente, y para liberar a los humanos de la toma de decisiones mundana, necesitamos trabajar de manera inteligente. Nuestra industria está enfrentando un inminente cambio generacional. Empleados experimentados pronto se retirarán y se irá con ellos décadas de experticia y conocimiento que son difíciles de ganar. Nosotros solo podemos decirles a nuestras nuevas contrataciones, "Hola, lea 300 terabytes de datos oscuros para que se pongan al día." Nosotros estamos yendo a confiar en nuevos enfoques para desarrollar, administrar y compartir la sabiduría-orientada-a-los-datos.

El despliegue de tecnologías analíticas para ayudar a iluminar conocimientos útiles no solo dentro de los datos brutos que ya están en nuestra posesión sino también en datos derivados representa una oportunidad de negocios potencialmente poderosa. Sin embargo, marcar sus esfuerzos de minería de datos y análisis e importar grandes almacenes de datos no-tradicionales provenientes de fuentes externas puede llevar a preguntas acerca de veracidad de los datos, integridad, legalidad, y lo apropiado de su uso. Esas son preguntas que pocas organizaciones pueden ignorar hoy.

Por la otra cara, el análisis profundo de más datos provenientes de una variedad de fuentes también puede arrojar señales que potencialmente podrían impulsar sus esfuerzos cibernéticos y de administración del riesgo. Además, la tendencia de las analíticas oscuras no se trata solo de desplegar herramientas analíticas crecientemente poderosas contra fuentes de datos no-aprovechados. Desde la perspectiva del riesgo cibernético, esta tendencia también se refiere a tener y usar esas y otras herramientas para inspeccionar tanto los datos en su posesión como datos de terceros que usted compre.

Cuando usted explore la tendencia de las analíticas oscuras, considere los siguientes problemas y oportunidades relacionados con el riesgo:

Fuentes de los datos: ¿En qué grado puede usted confiar en la integridad de los datos? Si usted no puede confirmar su exactitud, completitud, y consistencia, usted podría estar exponiendo a su compañía ante riesgos regulatorios, financieros, e incluso de marca. Lo mismo en relación con su autenticidad. ¿La fuente de los datos es quien dice que es? Si no, los datos podrían ser reciclados de otras fuentes dudosas o, peor aún, robados.

Respeto de la privacidad: El espectro de la ley de privacidad arroja una larga sombra sobre los datos de audio y video obtenidos fuera de la empresa. En muchos casos, las leyes de privacidad que aplican a los clips de audio o video están determinadas por la nacionalidad de los individuos que aparecen en ellos. De igual manera, en algunos países, incluso grabar una dirección del protocolo de Internet [Internet protocol (IP)] es considerada una violación de la privacidad. Al desarrollar la estrategia del riesgo cibernético de las analíticas oscuras, usted debe tener en mente los caprichos de la ley de seguridad. Finalmente, datos que pueden parecer benignos podrían, de hecho, transportar

potencial riesgo de seguridad si han sido derivados de analíticas. Por ejemplo, el análisis de datos del cliente puede sugerir una correlación entre preferencias de comidas del cliente y ciertas condiciones médicas o incluso su religión. Cuando usted comience a curar y analizar datos, ¿cómo coloca en funcionamiento los controles apropiados y administra los riesgos asociados de privacidad y legal? ¿Qué responsabilidad podría usted enfrentar si su archivo de datos contiene tales hallazgos correlativos?

Elaboración de modelos predictivos del riesgo: Cuando usted aplique analíticas a fuentes de datos no-tradicionales, puede haber oportunidades para crear modelos predictivos del riesgo que se basen en geografía, prácticas de contratación, préstamos, o diversos factores en el mercado. Esos modelos potencialmente podrían ayudarles a las compañías a desarrollar un entendimiento más matizado del sentimiento del empleado, del cliente, o del socio de negocios que, a su vez, pueda hacer posible desarrollar estrategias proactivas de mitigación del riesgo para abordar cada uno.

Proyección del panorama del riesgo de terceros: Las compañías globales pueden depender de cientos o incluso miles de vendedores que les proporcionan datos y otros servicios. Mediante analizar los datos provenientes de fuentes no-tradicionales, las compañías pueden ser capaces de crear modelos predictivos del riesgo que proporcionen perfiles más detallados del riesgo de sus proveedores. Algunos de los riesgos identificados probablemente pueden estar más allá de nuestro control – un punto a tener en mente cuando usted tenga como fuente datos de terceros.

Aprovechamiento de las señales de la red profunda: Los datos contenidos en esas partes de la red que actualmente están accesibles para los motores de búsqueda es demasiado enorme para que las organizaciones los aprovechen. Expandir ese universo ya-infinito para incluir datos no-tradicionales provenientes de la red profunda puede ofrecer oportunidades para el análisis de datos, pero desde el punto de vista del riesgo cibernético, también puede presentar considerables riesgos. La red oscura representa solo un componente pequeño de la red profunda más amplia, pero, una y otra vez, ha estado en la raíz de los desafíos y problemas cibernéticos.

“LA RED OSCURA
REPRESENTA SOLO UN
PEQUEÑO COMPONENTE
DE LA RED PROFUNDA
MÁS AMPLIA.”

Por lo tanto, probablemente amplificará los desafíos del riesgo y las complejidades que las compañías enfrentan si deciden explorarlo. Proceda con los ojos abiertos.

Dicho esto, mediante la aplicación de la modelación del riesgo a conjuntos de datos provenientes de la red

profunda más amplia, las organizaciones pueden ser capaces de ampliar adicionalmente su conocimiento en los reinos de seguridad cibernética, inteligencia competitiva, compromiso del cliente, y otras áreas de prioridad estratégica.

¿Dónde comienza usted?

Dentro de tres años, su organización puede encontrarse abrumada por inconmensurables volúmenes de datos no-estructurados generados por dispositivos del Internet de las Cosas. Trabajando hoy para desarrollar la disciplina y las herramientas que usted necesitará para administrar y hacer minería de todos estos datos oscuros puede ayudarle a su organización a generar hoy conocimientos orientados-a-los-datos al tiempo que se prepara para las aún mayores oportunidades que están por venir.

Este proceso comienza con una serie de pasos prácticos:

- **Haga las preguntas correctas:** Más que intentar descubrir e inventariar todos los datos oscuros ocultos dentro y fuera de su organización, trabaje con los equipos de negocio para identificar las preguntas específicas que quieran sean respondidas. Trabaje para identificar las fuentes potenciales de analíticas oscuras y las oportunidades no-descubiertas allí contenidas. Luego centre sus esfuerzos analíticos en las corrientes y fuentes de datos que sean particularmente relevantes. Por ejemplo, si mercadeo desea impulsar las ventas de equipos deportivos en cierta región, los equipos de analíticas pueden centrar sus esfuerzos en las corrientes de transacciones de venta en tiempo real, inventario, y datos sobre la fijación de precios de producto en almacenes seleccionados dentro de la región objetivo. Luego podrían complementar estos datos con datos históricos no-estructurados – análisis de videos en el almacén respecto del tráfico a pie, sentimiento social, comportamiento de quien influye, o incluso imágenes de pantallas o ubicación del producto a través de los sitios – para generar conocimientos más matizados.
- **Mire fuera de su organización:** Usted puede aumentar sus propios datos con información demográfica, de localizaciones, y estadística, públicamente disponible. No solo esto puede ayudarle a que sus equipos de analíticas generen reportes más expansivos, detallados – puede ubicar conocimientos en un contexto útil. Por ejemplo, un médico hace recomendaciones para un paciente de asma con base en su historia conocida de salud y un examen actual. Mediante revisar los datos locales del clima, también puede proporcionar soluciones de corto plazo para ayudarle a través de un brote durante una temporada de polen. En otro ejemplo, los empleados pueden analizar datos provenientes de herramientas geoespaciales, patrones de tráfico, y rotación de empleados a fin de determinar la extensión en la cual los niveles de satisfacción del empleado con el trabajo estén siendo adversamente impactados por los tiempos de viaje.
- **Aumente el talento de los datos:** Los científicos de datos son un recurso crecientemente valioso, especialmente quienes de manera ingeniosa pueden combinar modelación profunda y técnicas estadísticas con conocimientos específicos de la

industria o función y el enmarcar creativamente el problema. Avanzando, quienes tengan experticia demostrable en unas pocas áreas probablemente estarán en demanda. Por ejemplo, tanto el aprendizaje de máquina como el aprendizaje profundo requieren experticia programática – la capacidad para elaborar patrones establecidos para determinar la combinación apropiada de cuerpos de datos y el método para descubrir conocimientos razonables, defendibles. De igual manera, las habilidades de diseño gráfico y visual pueden crecientemente ser críticas dado que comunicar visualmente resultados y explicar las racionalidades son esenciales para la adopción organizacional más amplia. Finalmente, habilidades tradicionales tales como administración de datos maestros y arquitectura de datos serán tan valiosas como nunca antes – particularmente cuando más compañías comiencen a poner los fundamentos que necesitarán para las necesidades de datos diversos, expansivos, y en explosión del mañana.

- **Explore las herramientas avanzadas de visualización:** No todos en su organización serán capaces de digerir una impresión de estadísticas bayesianas avanzadas y aplicarlas a las prácticas de negocio. La mayoría de las personas necesitará entender el “y qué” y el “por qué” de los conocimientos analíticos complejos antes que puedan llevar el conocimiento a la acción. En muchas situaciones, la información puede ser más fácilmente diferida cuando sea presentada como una infografía, un tablero de mando, u otro tipo de representación visual. Los paquetes de software visual y de diseño pueden hacer más que gráficas llamativas tales como tablas de burbujas, nubes de palabras, mapas calientes – pueden impulsar la inteligencia del negocio mediante re-empaquetar grandes datos en cantidades más pequeñas, más significativas, entregando valor para los usuarios mucho más rápido. Adicionalmente, los conocimientos (y las herramientas) pueden ser hechos accesibles a través de la empresa, más allá del departamento de TI, y para usuarios de negocio en todos los niveles, para crear equipos trans-funcionales, más ágiles.
- **Véalo como un esfuerzo orientado-al-negocio:** Es tiempo de reconocer a las analíticas como una estrategia general del negocio más que como una función de TI. Para ello, trabaje con los colegas de la sala directiva para ganar apoyo para su enfoque de analíticas oscuras. Muchos CEO están haciendo de los datos la piedra angular de la estrategia general del negocio, lo cual obliga técnicas más sofisticadas y *accountability* por el manejo más deliberado de los activos subyacentes. Mediante entender la agenda y las metas de su organización, usted puede determinar el valor que tiene que ser entregado, definir las preguntas que se deben hacer, y decidir cómo aprovechar los datos disponibles para generar respuestas. Las analíticas de datos se convierten entonces en una ventaja orientada-al-conocimiento en el mercado. La mejor manera para ayudar a

asegurar que se compre esto es elaborar primero un piloto del proyecto que demostrará el ROI tangible que puede ser realizado por la organización con una estrategia de analíticas amplia para todo el negocio.

- **Piense de manera amplia:** Cuando usted desarrolle nuevas capacidades y estrategias, piense acerca de cómo usted puede extenderlas a través de la organización así como también a clientes, vendedores, y socios de negocio. Su nueva estrategia de datos se hace parte de su arquitectura de referencia que otros puedan usar.

La línea de resultados

Con la cada vez creciente cantidad de datos valiosos todavía inexplorados, la agregación, el análisis y el almacenamiento ya no son metas finales en la estrategia de analíticas de la organización ágil. Avanzando, los esfuerzos de las analíticas se centrarán en iluminar conocimientos poderosos de estrategia, cliente, y operacional ocultos en las fuentes de datos no-tradicionales y oscuras. Esté emocionado acerca del potencial de los datos no-estructurados y externos, pero manténgase conectado a tierra en las cuestiones específicas de negocio con alcance limitado y medible, y valor atribuible. Use esas preguntas para centrar sus esfuerzos relacionados con las analíticas de datos en las áreas que importan para su negocio – y evite perderse en lo desconocido crecientemente enorme.

AUTORES



TRACIE KAMBIES

Tracie Kambies es directivo en Deloitte Consulting LLP y el líder de administración de la información & analíticas de US retail y de IoT. Ofrece amplitud y profundidad de servicios para el mercado minorista en las áreas de estrategia de tecnología, analíticas y administración de la información, nube, y soluciones de IoT. Kambies es líder nacional de administración de la información & analíticas de aprendizaje, y tiene 16 años de experiencia en consultoría de negocios centrada en clientes en los sectores minorista, de consumo, y productos industriales.



NITIN MITTAL

Nitin Mittal es directive en Deloitte Consulting LLP y el líder de la práctica de US Analytics + Information Management. Ha asesorado clientes en su camino de analíticas y cómo podrían convertirse en organizaciones orientadas-al-conocimiento. El trabajo de Mittal se extiende desde consultoría en estrategia hasta implementación de ecosistemas analíticos para ayudarles a los clientes a aprovechar el potencial de las tecnologías exponenciales.



PAUL ROMA

Paul Roma lidera Deloitte Analytics, que ofrece poderosos resultados, no solo conocimientos, para algunos de los desafíos más difíciles que los negocios enfrentan hoy. Se especializa en desarrollo de estrategias de crecimiento e innovación, y en la aplicación de tecnologías disruptivas que les ayudan a darles a las compañías líderes del mundo a tener una ventaja orientada-al-conocimiento. La dedicación y el liderazgo de Roma han sido instrumentales en el avance de la práctica de Life Sciences and Health Care Technology, de Deloitte.



SANDEEP KUMAR SHARMA, PH.D.

Sandeep Sharma es el subdirector de tecnología jefe y director administrativo de la práctica de Analytics and Information Management de Deloitte LLP, con más de 18 años de experiencia global entregando inteligencia compleja de negocios, analíticas y programas de ciencia de datos para clientes en una variedad de industrias incluyendo servicios financieros, atención en salud, productos de consumo, telecomunicaciones, energía, y el sector público. Es arquitecto certificado TOGAF con una trayectoria de elaboración de soluciones estratégicas en analíticas predictivas, aprendizaje de máquina, BI social, grandes datos, analíticas en tiempo real, e información digital.

NOTAS FINALES

- ¹ Khalid Kark, Mark White, Bill Briggs, and Anjali Shaikh, *Navigating legacy: Charting the course to business value*, Deloitte University Press, November 10, 2016.
- ² Marc Goodman, "Most of the web is invisible to Google. Here's what it contains," *Popular Science*, April 1, 2015, www.popsci.com/dark-web-revealed.
- ³ Åse Dragland, "Big data, for better or worse: 90% of world's data generated over last two years," *Science Daily*, May 22, 2013, www.sciencedaily.com/releases/2013/05/130522085217.htm.
- ⁴ EMC Digital Universe with research and analysis by IDC, "The digital universe of opportunities: Rich data and the increasing value of the Internet of Things," April 2014, www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm.
- ⁵ Gartner Inc., "Gartner says 6.4 billion connected 'things' will be in use in 2016, up 30 percent from 2015," press release, November 10, 2015, www.gartner.com/newsroom/id/3165317.
- ⁶ Cisco, "Cisco Global Cloud Index: Forecast and methodology, 2014-2019," October 28, 2015, www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html.
- ⁷ Dan Vesset and David Schubmehl, "IDC FutureScape: Worldwide big data, business analytics, and cognitive software 2017 predictions," December 2016, International Data Corporation.
- ⁸ Ibid.
- ⁹ Seth Grymes, "Unstructured data and the 80 percent rule," *Breakthrough Analysis*, August 1, 2008, <https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>.
- ¹⁰ Cindy LaChapelle, "The cost of data management: Where is it headed in 2016?" *Datacenter Journal*, March 10, 2016, www.datacenterjournal.com/cost-data-storage-management-headed-2016/.
- ¹¹ Microsoft, "Multi-lingual, federated search solution provides global access to scientific research," March 24, 2015, <https://customers.microsoft.com/en-US/story/multilingual-federated-search-solution-provides-global>.
- ¹² David Barton, "Data mining in the deep web," *The Innovation Enterprise*, July 14, 2016, <https://channels.theinnovationenterprise.com/articles/data-mining-in-the-deep-web>.
- ¹³ Entrevista con Richard Chadderton, senior vice president of engagement and strategy, Indiana University Health, December 9, 2016.
- ¹⁴ Entrevista con Mark Lantzy, CIO of IU Health, November 21, 2016.
- ¹⁵ Jeff Simpson, Lokesh Ohri, and Kasey M. Lobaugh, *The new digital divide*, Deloitte University Press, September 12, 2016, <https://dupress.deloitte.com/dup-us-en/industry/retail-distribution/digital-divide-changing-consumerbehavior.html>.
- ¹⁶ Ibid.
- ¹⁷ Hilary Milnes, "How Stitch Fix's happy relationship with Pinterest helps customers," *Digiday*, March 16, 2016, <http://digiday.com/brands/how-stitch-fix-depends-on-pinterest-to-make-its-customers-happy/>.
- ¹⁸ Kim S. Nash, "Kroger tests sensors, analytics in interactive grocery shelves," *Wall Street Journal*, January 20, 2017, <http://blogs.wsj.com/cio/2017/01/20/kroger-tests-sensors-analytics-in-interactive-grocery-shelves/>.