

Una línea que conecta dos entidades que termina en dos marcas cortas designa una relación de uno a uno. Una línea que conecta dos entidades y termina con una pata de cuervo y una marca corta encima de ella indica una relación de uno a varios. La figura 6.11 muestra que un PEDIDO puede contener varios ARTICULO_LINEA. (Es posible ordenar una PIEZA muchas veces y que aparezca otras tantas como artículo de línea en un solo pedido.) Cada PIEZA solo puede tener un PROVEEDOR, pero muchos elementos PIEZA pueden ser proporcionados por el mismo PROVEEDOR.

No podemos enfatizarlo lo suficiente: si el modelo de datos de la empresa no es el correcto, el sistema no podrá dar buen servicio a la empresa. Los sistemas de la compañía no serán tan efectivos como podrían serlo debido a que tendrán que trabajar con datos que tal vez sean imprecisos, incompletos o difíciles de recuperar. Comprender los datos de la organización y la forma en que se deben representar en una base de datos es tal vez la lección más importante que usted puede aprender de este curso.

Por ejemplo, Famous Footwear, una cadena de zapaterías con más de 800 sucursales en 49 estados, no pudo lograr su objetivo de tener “el estilo correcto de zapato en la tienda apropiada para venderse al precio adecuado”, ya que su base de datos no estaba correctamente diseñada para ajustar con rapidez el inventario de las tiendas. La compañía tenía una base de datos relacional Oracle operando en una computadora de medio rango, pero el objetivo primordial para el que se diseñó la base de datos era producir informes estándar para la gerencia, en vez de reaccionar a los cambios en el mercado. La gerencia no pudo obtener datos precisos sobre artículos específicos en el inventario en cada una de sus tiendas. Para solucionar este problema, la compañía tuvo que crear una nueva base de datos en la que se pudieran organizar mejor los datos de las ventas y del inventario para realizar análisis y administrar el inventario.

6.3

¿CUÁLES SON LAS PRINCIPALES HERRAMIENTAS Y TECNOLOGÍAS PARA ACCEDER A LA INFORMACIÓN DE LAS BASES DE DATOS Y MEJORAR TANTO EL DESEMPEÑO DE NEGOCIOS COMO LA TOMA DE DECISIONES?

Las empresas utilizan sus bases de datos para llevar el registro de las transacciones básicas, como pagar a los proveedores, procesar pedidos, llevar el registro de los clientes y pagar a los empleados. Pero también se necesitan bases de datos para proveer información que ayude a la compañía a operar sus negocios con más eficiencia, y ayudar a los gerentes y empleados a tomar mejores decisiones. Si una compañía desea saber cuál producto es el más popular o quién es su cliente más rentable, la respuesta radica en los datos.

EL DESAFÍO DE BIG DATA

La mayoría de los datos recolectados por las organizaciones solían ser los datos de transacciones que podían caber fácilmente en filas y columnas de sistemas de administración de bases de datos relacionales. Ahora, somos testigos de una explosión de datos provenientes del tráfico Web, mensajes de correo electrónico y contenido de medios sociales (tweets, mensajes de estado), así como los datos generados por máquinas de los sensores (utilizados en medidores inteligentes, sensores de fabricación y medidores eléctricos) o de sistemas de e-commerce. Estos datos pueden ser estructurados o no estructurados y, por ende, tal vez no sean adecuados para productos de bases de datos relacionales que organicen los datos en forma de columnas y filas. Ahora usamos el término **big data** para describir estos conjuntos de datos con volúmenes tan grandes que están más allá de la capacidad de un DBMS común para capturar, almacenar y analizar.

Big Data no se refiere a una cantidad específica, sino por lo general a los datos en el rango de los petabytes y exabytes; es decir, de miles de millones a billones de registros,

todos de orígenes distintos. Los Big Data se producen en cantidades mucho mayores y con mucha más rapidez que los datos tradicionales. Por ejemplo, un solo motor de jet es capaz de generar 10 terabytes de datos en sólo 30 minutos, y hay más de 25,000 vuelos de aerolíneas a diario. Aun cuando los “tweets” se limitan a 140 caracteres cada uno, Twitter genera más de 8 terabytes de datos por día. De acuerdo con la empresa de investigación de tecnología International Data Center (IDC), los datos se duplican con creces cada dos años, por lo que la cantidad de datos disponibles para las organizaciones está aumentando en forma indiscriminada.

A las empresas les interesan los Big Data debido a que pueden revelar más patrones y anomalías interesantes que los conjuntos de datos más pequeños, con el potencial de proveer nuevas perspectivas en cuanto al comportamiento de los clientes, los patrones de clima, la actividad del mercado financiero u otros fenómenos. Sin embargo, para derivar un valor de negocios de estos datos, las organizaciones necesitan nuevas tecnologías y herramientas capaces de administrar y analizar datos no tradicionales junto con sus datos empresariales tradicionales.

INFRAESTRUCTURA DE INTELIGENCIA DE NEGOCIOS

Suponga que desea información concisa y confiable sobre las operaciones, tendencias y cambios actuales en toda la empresa. Si trabajara en una empresa de gran tamaño, tendría que reunir los datos necesarios de sistemas separados, como ventas, manufactura y contabilidad, e incluso desde fuentes externas, como los datos demográficos o de las competencias. Es probable que cada vez fuera más necesario usar Big Data. Una infraestructura contemporánea para la inteligencia de negocios tiene una variedad de herramientas para obtener información útil de todos los tipos diferentes de datos que usan las empresas en la actualidad, incluyendo Big Data semiestructurados y no estructurados en grandes cantidades. Estas capacidades incluyen almacenes de datos y mercados de datos, Hadoop, computación en memoria y plataformas analíticas. Algunas de estas capacidades están disponibles como servicios en la nube.

Almacenes de datos y mercados de datos

La herramienta tradicional para analizar datos corporativos durante las últimas dos décadas ha sido el almacén de datos. Un **almacén de datos** es una base de datos que almacena la información actual e histórica de interés potencial para los encargados de tomar decisiones en la compañía. Los datos se originan en muchos sistemas básicos de transacciones operacionales, como los sistemas de ventas, las cuentas de clientes, la manufactura, y pueden incluir datos de transacciones de sitios Web. El almacén de datos extrae los datos actuales e históricos de varios sistemas operacionales dentro de la organización. Estos datos se combinan con los datos de fuentes externas y se transforman al corregir los datos imprecisos e incompletos y reestructurar los datos para generar informes gerenciales y realizar análisis antes de cargarlos en el almacén de datos.

El almacén de datos pone los datos a disposición de todos según sea necesario, pero no se puede alterar. Un sistema de almacén de datos también provee un rango de herramientas de consulta ad hoc y estandarizadas, herramientas analíticas y facilidades de informes gráficos.

A menudo las empresas crean almacenes de datos a nivel empresarial, donde un almacén de datos central da servicio a toda la organización, o crean almacenes de datos más pequeños y descentralizados conocidos como mercados de datos. Un **mercado de datos** es un subconjunto de un almacén de datos, en el cual se coloca una porción sintetizada o con alto grado de enfoque en los datos de la organización en una base de datos separada para una población específica de usuarios. Por ejemplo, una compañía podría desarrollar mercados de datos sobre marketing y ventas para lidiar con la información de los clientes. El vendedor de libros Barnes & Noble solía mantener una serie de mercados de datos: uno para los datos sobre los puntos de venta en las tiendas minoristas, otro para las ventas de las librerías universitarias y un tercero para las ventas en línea.

Hadoop

Los productos de DBMS relacionales y almacenes de datos no se adaptan bien para organizar y analizar Big Data o datos que no caben fácilmente en las columnas y filas utilizadas en sus modelos de datos. Para manejar datos no estructurados y semiestructurados en grandes cantidades, así como datos estructurados, las organizaciones usan **Hadoop**, que es un marco de trabajo de software de código abierto, administrado por la Fundación de Software Apache, lo que permite el procesamiento paralelo distribuido de enormes cantidades de datos a través de computadoras económicas. Descompone un problema de Big Data en varios subproblemas, los distribuye entre miles de nodos de procesamiento de computadoras económicas y luego combina el resultado en un conjunto de datos de menor tamaño que es más fácil de analizar. Tal vez usted ya haya usado Hadoop para encontrar la mejor tarifa aérea en Internet, obtener indicaciones para llegar a un restaurante, realizar una búsqueda en Google o conectarse con un amigo en Facebook.

Hadoop consta de varios servicios clave: el sistema de archivos distribuidos Hadoop (HDFS) para almacenamiento de datos y MapReduce para procesamiento de datos en paralelo de alto rendimiento. HDFS enlaza entre sí los sistemas de archivos en los numerosos nodos en un clúster Hadoop para convertirlos en un gran sistema de archivos. MapReduce de Hadoop se inspiró en el sistema MapReduce de Google para desglosar el procesamiento de enormes conjuntos de datos y asignar trabajo a los diversos nodos en un clúster. HBase, la base de datos no relacional de Hadoop, ofrece un acceso rápido a los datos almacenados en HDFS y una plataforma transaccional para ejecutar aplicaciones en tiempo real de alta escala.

Hadoop puede procesar grandes cantidades de cualquier tipo de datos, incluyendo datos transaccionales estructurados, datos poco estructurados como las fuentes de Facebook y Twitter, datos complejos como los archivos de registro de servidor Web y datos de audio y video no estructurados. Hadoop se ejecuta en un clúster de servidores económicos y pueden agregarse o eliminarse procesadores según sea necesario. Las empresas usan Hadoop para analizar volúmenes muy grandes de datos, así como para un área de concentración para datos no estructurados y semiestructurados antes de cargarlos en un almacén de datos. Facebook almacena gran parte de sus datos en un enorme clúster Hadoop, que contiene cerca de 100 petabytes, alrededor de 10,000 veces más información que la Biblioteca del Congreso estadounidense. Yahoo usa Hadoop para rastrear el comportamiento de los usuarios de modo que pueda modificar su página de inicio y adaptarla a sus intereses. La empresa de investigación de ciencias de la vida NextBio usa Hadoop y HBase para procesar datos para empresas farmacéuticas que realizan investigación genómica. Los principales distribuidores de bases de datos como IBM, Hewlett-Packard, Oracle y Microsoft tienen sus propias distribuciones de software de Hadoop. Otros distribuidores ofrecen herramientas para meter y sacar datos de Hadoop, o para analizarlos dentro de Hadoop.

Computación en memoria

Otra forma de facilitar el análisis de Big Data es utilizar la **computación en memoria**, que depende principalmente de la memoria principal (RAM) de la computadora para el almacenamiento de datos (los DBMS convencionales usan sistemas de almacenamiento de datos). Los usuarios acceden a los datos almacenados en la memoria principal del sistema, con lo cual se eliminan los cuellos de botella por los procesos de recuperación y lectura de datos en una base de datos tradicional basada en discos, y se reducen de manera drástica los tiempos de respuesta de las consultas. El procesamiento en memoria hace posible que conjuntos muy grandes de datos, del tamaño de un mercado de datos o de un almacén pequeño de datos, residan totalmente en la memoria. Los cálculos de negocios complejos que solían tardar horas o días pueden completarse en cuestión de segundos, y esto puede lograrse incluso en dispositivos portátiles (vea la Sesión interactiva: tecnología).

El capítulo anterior describe algunos de los avances en la tecnología de hardware de computadora contemporánea que hacen posible el procesamiento en memoria, como los poderosos procesadores de alta velocidad, el procesamiento multinúcleo y los precios cada vez menores de la memoria de computadora. Estas tecnologías ayudan a las empresas a optimizar el uso de la memoria y aceleran el rendimiento del procesamiento, a la vez que reducen los costos.

nes y diseñar un programa para esa consulta específica, un proceso que hubiera tomado 36 horas.

Ahora, usando HANA, ARI puede extraer rápidamente sus amplios recursos de datos y generar predicciones con base en los resultados. Por ejemplo, la empresa puede producir cifras precisas sobre los costos de operar una flotilla de cierto tamaño a través de determinada ruta en industrias específicas durante cierto tipo de clima y predecir el impacto de los cambios en alguna de esas variables. Y puede hacerlo casi con tanta facilidad como la de proveer a sus clientes un historial simple de sus gastos de combustible. Con esta información tan útil ARI provee más valor a sus clientes.

HANA también redujo el tiempo requerido para cada transacción manejada por los call centers de ARI (desde el momento en que un miembro del personal del call center toma una llamada hasta la recuperación y entrega de la información solicitada) en un 5%. Como el personal de los

call centers representa el 40% de la sobrecarga directa de ARI, esa reducción en tiempo se traduce en grandes ahorros en costo.

ARI planea tener algunas de estas capacidades de generación de informes y análisis en tiempo real disponibles en dispositivos móviles, lo cual permitirá a los clientes aprobar al instante varios procedimientos operacionales, como la autorización de reparaciones de mantenimiento. Los clientes también podrán usar las herramientas móviles para una perspectiva instantánea de las operaciones de sus flotillas, con un nivel de detalle como el historial de los neumáticos de un vehículo específico.

Fuentes: "Driving 2 Million Vehicles with SAP Data", www.sap.com, visitado el 1 de febrero de 2014; www.arifleet.com, visitado el 1 de febrero de 2014, y "ARI Fleet Management Drives Real-Time Analytics to Customers", *SAP InsiderPROFILES*, 1 de abril de 2013.

PREGUNTAS DEL CASO DE ESTUDIO

1. ¿Por qué era tan problemática la administración de datos en ARI?
2. Describa las capacidades anteriores de ARI en cuanto a análisis de datos y generación de informes, y su impacto en el negocio.
3. ¿Fue SAP HANA una buena solución para ARI? ¿Por qué?
4. Describa los cambios en los negocios como resultado de adoptar HANA.

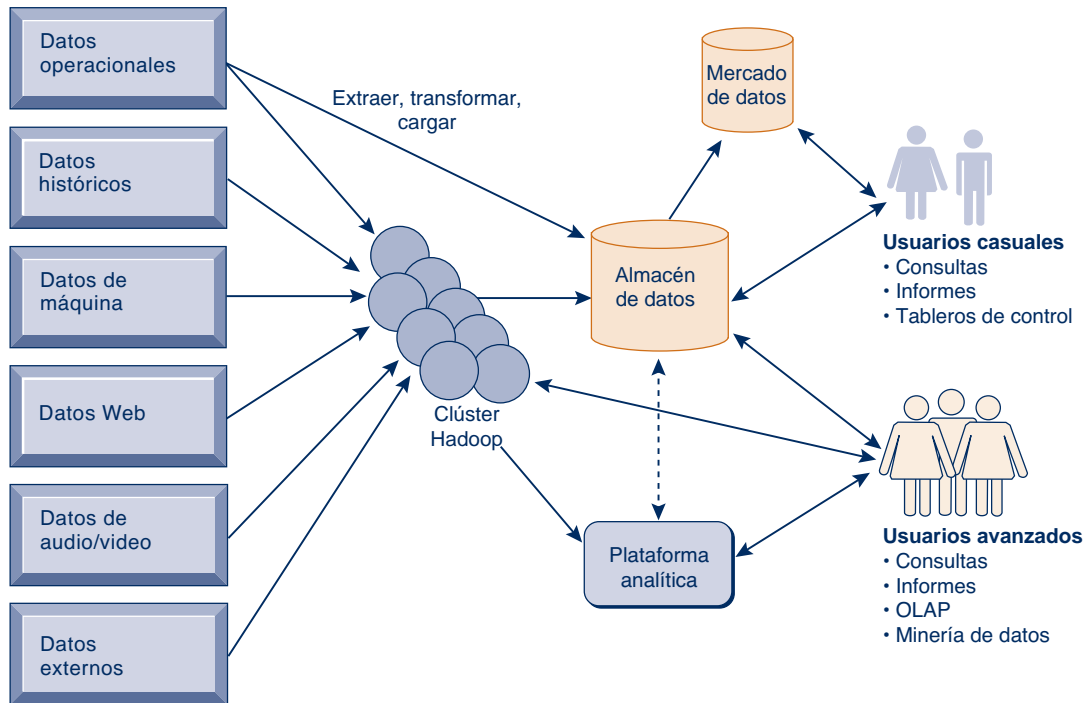
Los principales productos comerciales para la computación en memoria son: High Performance Analytics Appliance (HANA) de SAP, y Oracle Exalytics. Cada uno ofrece un conjunto de componentes de software integrados, incluyendo software de base de datos en memoria y software de análisis especializado, que se ejecutan en hardware optimizado para el trabajo de cómputo en memoria.

Plataformas analíticas

Los distribuidores de bases de datos comerciales han desarrollado **plataformas analíticas** especializadas de alta velocidad que utilizan tecnología tanto relacional como no relacional y están optimizadas para analizar conjuntos de datos de gran tamaño. Las plataformas analíticas como IBM Netezza y Oracle Exadata cuentan con sistemas de hardware-software preconfigurados que están diseñados de manera específica para el procesamiento de consulta y los análisis. Por ejemplo, IBM Netezza tiene componentes de base de datos, servidor y almacenamiento estrechamente integrados que manejan consultas analíticas complejas 10 a 100 veces más rápido que los sistemas tradicionales. Las plataformas analíticas también incluyen sistemas en memoria y sistemas de administración de bases de datos no relacionales. Ahora, las plataformas analíticas están disponibles como servicios en la nube.

La figura 6.12 ilustra una infraestructura de inteligencia de negocios contemporánea que usa las tecnologías que acabamos de describir. Los datos actuales e históricos se extraen de varios sistemas operacionales junto con datos Web, datos generados por máquinas, datos de audio/visuales no estructurados y datos provenientes de fuentes externas, que se han reestructurado y organizado para generación de informes y análisis. Los clústeres Hadoop preprocesan los Big Data para usarlos en el almacén de datos, mercados de datos o en una plataforma analítica, o para que los usuarios avanzados los

FIGURA 6.12 INFRAESTRUCTURA CONTEMPORÁNEA DE INTELIGENCIA DE NEGOCIOS



Una infraestructura contemporánea de inteligencia de negocios cuenta con capacidades y herramientas para administrar y analizar grandes cantidades y distintos tipos de datos provenientes de varias fuentes. Se incluyen herramientas de consulta y generación de informes fáciles de usar para los usuarios de negocios casuales y conjuntos de herramientas analíticas más sofisticadas para usuarios avanzados.

consulten de manera directa. Los resultados incluyen informes y tableros de control, así como resultados de las consultas. En el capítulo 12 veremos con mayor detalle los diversos tipos de usuarios BI y generación de informes BI.

HERRAMIENTAS ANALÍTICAS: RELACIONES, PATRONES, TENDENCIAS

Una vez que los datos se capturan y organizan mediante el uso de las herramientas para inteligencia de negocios que acabamos de describir, están disponibles para un posterior análisis utilizando el software para consultas e informes de bases de datos, el análisis de datos multidimensional (OLAP) y la minería de datos. En esta sección le presentaremos estas herramientas; en el capítulo 12 veremos más detalles sobre el análisis de inteligencia de negocios y aplicaciones.

Procesamiento analítico en línea (OLAP)

Suponga que su compañía vende cuatro productos distintos: tuercas, pernos, arandelas y tornillos en las regiones Este, Oeste y Central. Si deseara hacer una pregunta muy directa, por ejemplo, cuántas arandelas se vendieron durante el trimestre pasado, podría encontrar la respuesta con facilidad al consultar su base de datos de ventas. Pero ¿qué pasaría si quisiera saber cuántas arandelas se vendieron en cada una de sus regiones de ventas, para comparar los resultados actuales con las ventas proyectadas?

Para obtener la respuesta, necesitaría el **procesamiento analítico en línea (OLAP)**. OLAP soporta el análisis de datos multidimensional, el cual permite a los usuarios ver los

misimos datos de distintas formas mediante el uso de varias dimensiones. Cada aspecto de información —producto, precios, costo, región o periodo de tiempo— representa una dimensión distinta. Así, un gerente de productos podría usar una herramienta de análisis de datos multidimensional para saber cuántas arandelas se vendieron en el Este en junio, cómo se compara esa cifra con la del mes anterior y con la de junio del año anterior, y cómo se compara con el pronóstico de ventas. OLAP permite a los usuarios obtener respuestas en línea a preguntas ad hoc como éstas en un tiempo muy corto, incluso cuando los datos se almacenan en bases de datos muy grandes, como las cifras de ventas de varios años.

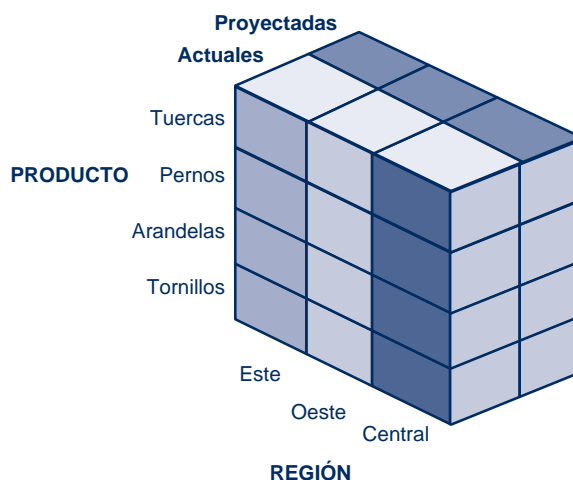
La figura 6.13 muestra un modelo multidimensional que podría crearse para representar productos, regiones, ventas reales y ventas proyectadas. Una matriz de ventas actuales se puede apilar encima de una matriz de ventas proyectadas para formar un cubo con seis caras. Si gira el cubo 90° en un sentido, la cara que se muestre será la del producto contra ventas actuales y proyectadas; si lo gira de nuevo 90°, verá la cara de la región contra ventas actuales y proyectadas, y si lo gira 180° a partir de la vista original, verá las ventas proyectadas y producto contra región. Se pueden anidar cubos dentro de otros cubos para crear vistas complejas de datos. Una compañía podría utilizar una base de datos multidimensional especializada, o una herramienta que cree vistas multidimensionales de datos en las bases de datos relacionales.

Minería de datos

Las consultas en las bases de datos tradicionales responden a preguntas como: “¿cuántas unidades del producto número 403 se enviaron en febrero de 2013?” El OLAP (análisis multidimensional) soporta solicitudes mucho más complejas de información, como: “comparar las ventas del producto 403 relativas con el plan por trimestre y la región de ventas durante los últimos dos años”. Con OLAP y el análisis de datos orientados a consultas, los usuarios necesitan tener una buena idea sobre la información que están buscando.

La **minería de datos** está más orientada al descubrimiento, ya que provee perspectivas hacia los datos corporativos que no se pueden obtener mediante OLAP, al encontrar patrones y relaciones ocultas en las bases de datos grandes e inferir reglas a partir de estos patrones y relaciones, para predecir el comportamiento a futuro. Los patrones y

FIGURA 6.13 MODELO DE DATOS MULTIDIMENSIONAL



La vista que se muestra es la de producto contra región. Si gira el cubo 90 grados, la cara mostrará la vista de producto contra las ventas actuales y proyectadas; si lo gira 90 grados otra vez, verá la vista de región contra ventas actuales y proyectadas. Es posible obtener otras vistas.

reglas se utilizan para guiar la toma de decisiones y pronosticar el efecto de esas decisiones. Los tipos de información que se pueden obtener de la minería de datos son: asociaciones, secuencias, clasificaciones, agrupamientos y pronósticos.

- Las *asociaciones* son ocurrencias vinculadas a un solo evento. Por ejemplo, un estudio de los patrones de compra en supermercados podría revelar que cuando se compran frituras de maíz, el 65% de veces se compra un refresco de cola, pero cuando hay una promoción, es el 85% de veces. Esta información ayuda a los gerentes a tomar mejores decisiones debido a que descubren la rentabilidad de una promoción.
- En las *secuencias*, los eventos se vinculan en el transcurso del tiempo. Por ejemplo, podríamos descubrir que si se compra una casa, el 65% de veces se compra un refrigerador nuevo dentro de las siguientes dos semanas, y el 45% se compra un horno dentro del mes posterior a la compra de la casa.
- La *clasificación* reconoce los patrones que describen el grupo al que pertenece un elemento, para lo cual se examinan los elementos existentes que hayan sido clasificados y se infiere un conjunto de reglas. Por ejemplo, las empresas, como las compañías de tarjetas de crédito o las telefónicas, se preocupan por la pérdida de clientes estables. La clasificación ayuda a descubrir las características de los clientes con probabilidades de dejar de serlo y puede proveer un modelo para ayudar a los gerentes a predecir quiénes son esos clientes, de modo que puedan idear campañas especiales para retenerlos.
- El *agrupamiento* funciona de una manera similar a la clasificación cuando aún no se han definido grupos. Una herramienta de minería de datos puede descubrir distintas agrupaciones dentro de los datos, como el hecho de encontrar grupos de afinidad para tarjetas bancarias o particionar una base de datos en grupos de clientes con base en la demografía y los tipos de inversiones personales.
- Aunque estas aplicaciones implican predicciones, el *pronóstico* utiliza las predicciones de una manera distinta. Se basa en una serie de valores existentes para pronosticar cuáles serán los otros valores. Por ejemplo, el pronóstico podría encontrar patrones en los datos para ayudar a los gerentes a estimar el futuro valor de variables continuas, como las cifras de ventas.

Estos sistemas realizan análisis de alto nivel de los patrones o tendencias, pero también pueden profundizar para proveer más detalles cuando sean necesarios. Hay aplicaciones de minería de datos para todas las áreas funcionales de negocios, y también para el trabajo gubernamental y científico. Un uso popular de la minería de datos es el de proveer análisis detallados de los patrones en los datos de los consumidores para las campañas de marketing de uno a uno, o para identificar a clientes rentables.

Entertainment, anteriormente conocida como Harrah's Entertainment, es la segunda compañía de apuestas más grande del mundo. Analiza continuamente los datos sobre sus clientes que se recopilan cuando las personas juegan en las máquinas tragamonedas o utilizan sus casinos y hoteles. El departamento de marketing corporativo utiliza esta información para crear un perfil de apuestas detallado, con base en el valor continuo de un cliente específico para la compañía. Por ejemplo, la minería de datos permite a Caesars conocer la experiencia de juego favorita de un cliente regular en uno de sus casinos en los barcos, junto con las preferencias de esa persona en cuanto al alojamiento, los restaurantes y el entretenimiento. Esta información guía las decisiones gerenciales sobre cómo cultivar los clientes más rentables y animarlos a que gasten más, y también sobre cómo atraer más clientes con un alto potencial de generación de ingresos. La inteligencia de negocios mejoró tanto las ganancias de Caesars que se convirtió en la pieza central de la estrategia de negocios de la empresa.

Minería de texto y minería Web

Se cree que los datos no estructurados, que en su mayoría están organizados en forma de archivos de texto, representan más del 80% de la información útil de una organización y son una de las principales fuentes de Big Data que las empresas desean analizar. El correo

electrónico, los memorándums, las transcripciones de los call centers, las respuestas a las encuestas, los casos legales, las descripciones de patentes y los informes de servicio son todos elementos valiosos para encontrar patrones y tendencias que ayuden a los empleados a tomar mejores decisiones de negocios. En la actualidad hay herramientas de **minería de texto** disponibles para ayudar a las empresas a analizar estos datos. Estas herramientas pueden extraer elementos clave de los conjuntos de datos extensos no estructurados, descubrir patrones y relaciones, así como sintetizar la información.

Las empresas podrían recurrir a la minería de texto para analizar las transcripciones de los call center de servicio al cliente para identificar las principales cuestiones de servicio y reparación, o para medir el sentimiento de los clientes con respecto a su empresa. El software de análisis de opiniones es capaz de extraer los comentarios de texto en un mensaje de correo electrónico, blog, conversación de social media o formulario de encuesta para detectar las opiniones favorables y desfavorables sobre temas específicos.

Por ejemplo, el corredor de saldos Charles Schwab usa el software Attensity Analyze para analizar cientos de miles de interacciones de sus clientes cada mes. El software analiza las notas de servicio de los clientes de Schwab, los correos electrónicos, las respuestas de las encuestas y las discusiones en línea para descubrir señales de descontento que puedan provocar que un cliente deje de usar los servicios de la empresa. Attensity puede identificar automáticamente las diversas “voces” que usan los clientes para expresar su retroalimentación (como una voz positiva, negativa o condicional) para señalar la intención de una persona de comprar, su intención de abandonar, o la reacción a un producto o mensaje de marketing específico. Schwab usa esta información para tomar acciones correctivas como establecer una comunicación directa del corredor con el cliente y tratar de resolver con rapidez los problemas que lo tienen descontento.

Web es otra fuente de datos extensos no estructurados para revelar patrones, tendencias y perspectivas en relación con el comportamiento de los clientes. El descubrimiento y análisis de los patrones útiles y la información proveniente de World Wide Web se denominan minería Web. Las empresas podrían recurrir a la minería Web para que les ayude a comprender el comportamiento de los clientes, evaluar la efectividad de un sitio Web específico o cuantificar el éxito de una campaña de marketing. Por ejemplo, los comerciantes utilizan los servicios Google Trends y Google Insights for Search, que rastrean la popularidad de varias palabras y frases utilizadas en las consultas de búsqueda de Google para saber en qué están interesadas las personas y qué les interesa comprar.

La minería Web busca patrones en los datos a través de la minería de contenido, la minería de estructura y la minería de uso. La minería de contenido Web es el proceso de extraer conocimiento del contenido de páginas Web, lo cual puede incluir datos de texto, imágenes, audio y video. La minería de estructura Web examina los datos relacionados con la estructura de un sitio Web específico. Por ejemplo, los vínculos que apuntan a un documento indican su popularidad, en tanto que los que salen de un documento indican la riqueza, o tal vez la variedad de temas cubiertos en él. La minería de uso Web examina los datos de interacción de los usuarios registrados por un servidor Web cada vez que se reciben solicitudes relacionadas con los recursos de un sitio Web. Los datos de uso registran el comportamiento del usuario cuando navega o realiza transacciones en el sitio Web y recolecta los datos en un registro del servidor. Al analizar esos datos, las compañías pueden determinar el valor de ciertos clientes específicos, las estrategias de marketing cruzado entre los diversos productos y la efectividad de las campañas promocionales.

El caso al final del capítulo describe las experiencias de las organizaciones al usar las herramientas analíticas y las tecnologías de inteligencia de negocios que hemos descrito para lidiar con los desafíos de los “big data”.

LAS BASES DE DATOS Y WEB

¿Alguna vez ha tratado de usar la Web para realizar un pedido o ver un catálogo de productos? Si su respuesta es positiva, es probable que haya usado un sitio Web vinculado