

Conceptos básicos de Big Data

Probablemente ha oído hablar de big data, ¿aunque qué es exactamente y cómo puede utilizarse para proporcionar a su organización una ventaja competitiva? En este curso, va a aprender las características clave de big data, los beneficios que puede proporcionar y los tipos de datos que incorpora. También conocerá diferentes alternativas de aplicación, en función del sector de mercado. Este curso también incluye algunos desafíos fundamentales para la seguridad de big data y algunas mejores prácticas para gestionar big data a través de un ciclo de vida efectivo de la información.

Tabla de contenido

1. [Conceptos básicos de Big Data](#)
2. [Características de Big Data](#)
3. [Beneficios de Big Data](#)
4. [Tipos de fuentes de datos para Big Data](#)
5. [Big Data por sector de mercado](#)
6. [Big Data y la seguridad](#)
7. [Fases del ciclo de vida de los datos](#)

Conceptos básicos de Big Data

[Título del curso: Conceptos básicos de Big Data] Probablemente haya oído hablar de big data. ¿Aunque conoce exactamente qué es? O cómo puede utilizarse para proporcionar una ventaja competitiva. En este curso va a aprender las características clave de big data, sus beneficios y los tipos de datos que cubre. También conocerá diferentes alternativas de aplicación, en función del sector de mercado. Este curso también incluye algunos retos de la seguridad de big data y examina un ciclo de vida de datos típico.

Características de Big Data

[Título del tema: Características de Big Data] Big data. ¿Qué es exactamente? Piense en big data como en datos internos y externos recopilados de fuentes tradicionales y digitales. Se analizan para impulsar mejoras en su negocio, lo que a veces conduce a descubrimientos inesperados.

A veces los datos ya están estructurados de una cierta forma, aunque, por lo general, están parcialmente o nada estructurados; piense en metadatos o en Twitter. Los datos siempre están presentes, aunque hoy se procesan en un abrir y cerrar de ojos.

Es probable que haya oído hablar de las tres V de big data; alto volumen, alta velocidad y amplia variedad de datos.

La palabra "big" con respecto a los datos representa el volumen. Aunque los volúmenes gigantes de datos no es algo nuevo, lo que ha cambiado es el deseo de interpretar dicha información.

El problema es que los sistemas de gestión de bases de datos relacionales Tradicionales,

RDBMS, no pueden manejar dichos volúmenes. Están plagados de largos tiempos de consulta y datos sin estructurar. También son caros y a veces poco fiables.

Otro aspecto del volumen es cuánto tiempo siguen siendo útiles los datos almacenados, ¿cuándo los elimina? Los datos antiguos ocupan espacio y pasan a ser erróneos rápidamente

Otro problema es la alta velocidad de los datos. Es la velocidad de los datos de entrada. Imagine un grifo abierto al máximo. ¿Puede obtener toda el agua? ¿Puede hacerlo de una vez sin pausa? Volviendo a los datos, incluso cuando se capturan datos a alta velocidad, se tiene que estar seguro de su consistencia e integridad. La rapidez, o latencia, es un aspecto de la velocidad; es si los datos se están capturando a una velocidad razonable, o muy lentamente, lo que puede hacer que los datos dejen de ser útiles.

Para que los análisis reales tengan valor, el almacenamiento y recuperación de datos tienen que realizarse a alta velocidad. Por ejemplo, piense sobre su sitio de streaming de música favorita. Cada vez que inicie sesión, verá una lista nueva de canciones y álbumes, generada a partir de todos los datos recopilados sobre usted.

Todos esos datos son muy difíciles de gestionar. Intentar hacer coincidir caudales de datos con eventos específicos es difícil, más incluso por la gran variedad de datos. Vídeos, fotos, música, GPS y señales de sensores, y miles de millones de mensajes de texto de redes sociales. No están estructurados, y se estima que esos datos suponen el 80% de lo que se almacena.

Existen varios tipos de datos, aunque también una variedad de recursos, internos y externos a su empresa. Un punto de inflexión real será encontrar nuevas formas de almacenar y recuperar datos lo antes posible. O formas de empaquetarlos para mejorar el análisis. Mientras que volumen, velocidad y variedad son las grandes V, no olvide que también hay algunas V secundarias. Valor, veracidad, variabilidad, viscosidad y vitalidad; todas son relevantes para cualquier plan de big data.

Big data es un término en evolución, en una industria que evoluciona. Se esperan actualizaciones a medida que cambien nuestra comprensión de qué son los datos, y qué podrían ser.

Beneficios de Big Data

[Título del tema: Beneficios de Big Data] Los consumidores de hoy son muy exigentes. Es fácil investigar un producto antes de comprarlo; podría incluso hacerlo en la tienda. Con big data, puede trazar un perfil de sus clientes y captar sus gustos personales en tiempo real. Vamos a examinar algunos otros beneficios de big data.

Las reducciones de costes y tiempo elevarán los beneficios casi siempre. Hadoop, y tecnologías similares de big data, ofrecen grandes mejoras frente la tecnología antigua, como almacenes de datos. Existe una tendencia en la que grandes empresas como Citi combinan la nueva y la antigua tecnología para reducir costes. Piense sobre esto: puede modelar y probar millones de variaciones sobre un producto, y simularlas instantáneamente. El ahorro en tiempo, dinero y esfuerzo es realmente alto. Los análisis de big data se centran en la eficiencia de sus procesos de negocio.

Por ejemplo, en lugar de sustituir todas las máquinas de una fábrica en un intervalo de rotura promedio, big data permite comprobaciones específicas en tiempo real para identificar máquinas específicas, o incluso componentes individuales, y sustituirlas cuando

sea necesario. Menos tiempo de parada, menos dinero gastado, más efectividad. El tiempo es dinero, y la nueva tecnología acelera la toma de decisiones. Hadoop ofrece análisis en memoria rapidísimos.

Caesars, una empresa de juegos, adoptó Hadoop. Siempre recopilaban y analizaban información de sus máquinas de juego, programa de fidelidad y clickstreams web. Aunque big data permite analizarlos en tiempo real. Ahora pueden reaccionar instantáneamente mientras que el cliente sigue en el casino. Por ejemplo, si alguien está teniendo muy mala suerte en las máquinas de juego, un empleado de Caesars podría acercarse con un cupón para tentarle a continuar.

Además de mejorar los servicios existentes, big data genera muchas nuevas oportunidades. Por primera vez, las empresas offline, como las empresas online han hecho durante años, intentan crear nuevos productos y servicios con big data.

Los equipos deportivos usan datos sobre quién asistió a sus partidos y quién viajó hasta allí. Además de qué actividades extra hacen los seguidores antes y después del partido, como qué cadena de comida es la más popular. Este tipo de datos es inestimable para que la publicidad dirigida llegue a los clientes. Otro ejemplo es analizar posts de redes sociales. Esto permite conocer qué sienten las personas sobre su marca y calidad del servicio.

Otro desarrollo consiste en que las empresas pueden vender sus perspectivas de mercado como datos de tendencias, después de eliminar toda la información personal. Los propios datos pueden ser un caudal de ingresos valioso.

Aunque no solo son los únicos beneficios. Big data abre el diálogo con clientes individuales. Puede retocar sitios web sobre la marcha. Los análisis de riesgos son más fáciles, y sus datos están más seguros, con iniciativas como ciudades inteligentes y atención médica personalizada.

Es probable que su organización ya esté recopilando big data. Aprovechélo adecuadamente, y estará en primera línea para recoger sus frutos.

Tipos de fuentes de datos para Big Data

[Título del tema: Tipos de fuentes de datos para Big Data] ¿Sus datos se parecen más a un archivador financiero organizado perfectamente o a una buhardilla polvorienta llena de objetos? Bueno, no es tan diferente de cómo ven las bases de datos los datos estructurados y sin estructurar.

Los datos financieros, como números, fechas, direcciones o nombres de las cosas, son unidades de datos denominadas cadenas. Tienen longitud y formato. Esto hace que sean estructurados. Es fácil almacenar datos estructurados en una base de datos. Puede usar un lenguaje de consulta estructurado, o SQL, para extraer información útil y específica. Los recursos típicos incluyen datos de planificación de gestión empresarial, sistemas de gestión de relaciones con el cliente o sus datos financieros.

Algunos datos estructurados son generados por máquinas, como datos de punto de venta. Sistema de posicionamiento global, GPS, Identificación de radiofrecuencia, RFID, y otros datos de sensor, son creados sin intervención humana, como lo son estadísticas de servidores, aplicaciones y redes.

El otro tipo de datos estructurados resulta de la interacción humana con una máquina. Los datos capturados de formularios a mano son un ejemplo, ya que son clickstreams. Casinos podría monitorizar todos los movimientos en una de sus máquinas de juego. Si intenta

predecir patrones de uso, los datos estructurados son su mejor amigo. No obstante, solo el 20% de los datos está estructurado. Aunque piense en YouTubers que cargan 48 horas de vídeo al minuto. Esto supone un caudal masivo de datos sin formato específico, por lo que son datos sin estructurar. Los datos sin estructurar constituyen el 80% restante de big data. Es la gran mayoría.

Considere todas las imágenes y vídeos capturados cada día. ¿Cuántos datos son? Piense solo en la cantidad de datos dejados en Google Earth. Esos ejemplos son todos generados con máquinas. Datos científicos de la atmósfera, radar y sonar proceden también de máquinas. Aunque aplique alguna interacción humana a la mezcla y obtendrá todo el texto en documentos, PDF y correos electrónicos; obtendrá cada mínimo detalle de las entradas en redes sociales; mensajes de texto de móviles; y el contenido de sitios web como YouTube, Flickr e Instagram.

Ahora podría estar pensando: no estoy de acuerdo, muchos de esos ejemplos tienen estructura. Si fuera así, no es el único. Algunas personas consideran que en un correo electrónico hay una estructura definida, es un correo electrónico. El problema es que las herramientas de minería de datos no pueden obtener ninguna información útil de él. ¿Cómo se determina el contenido del correo? ¿Qué incluye? ¿De qué trata?

A pesar del aparente número incontable de tipos de datos sin estructurar, todos tienen una cosa en común. Su formato no importa. Pueden almacenarse sin que el sistema conozca qué son.

El reto de muchas organizaciones ya no es manejar toda esa información, el reto es hacer que sea significativa. Mientras más entienda y haga que los datos sin estructurar sean útiles, mayor valor ofrecerá a su empresa.

Big Data por sector de mercado

[Título del tema: Big Data por sector de mercado] ¿Cómo puede big data ayudarle en su industria? Prácticamente cualquier organización en cualquier segmento de mercado puede usar big data. Hay para todos.

Existen usos evidentes para banca y finanzas. Una abundancia de información significa que los productos pueden adaptarse para satisfacer al cliente. Seguir el comportamiento ofrece perspectivas sobre cómo monetizar hábitos de compra. Aunque con todo este caudal de datos, mantenerlos seguros es un reto continuo. Las instituciones bancarias y financieras deben estar alerta y con sistemas que minimicen el riesgo y eviten fraudes.

El GE Capital Banca Minorista, por ejemplo, ofrece ahora tarjetas de crédito para satisfacer necesidades del cliente en función de los datos recopilados. Están vinculados con una iniciativa que recopila datos del gasto del cliente.

Se sabe que el gobierno de EE.UU. posee seis de los diez superequipos más importantes del mundo. Big data tiene muchos usos en el sector gubernamental. Es mucho más fácil para el seguimiento, gestión y optimización de servicios. El uso de análisis permite a big data ayudar a las agencias gubernamentales. Los delitos pueden reducirse o evitarse. Y la congestión de tráfico puede analizarse y aliviarse.

Otro ejemplo, la Administración de Alimentos y Medicamentos utiliza big data para estudiar patrones en enfermedades alimentarias, lo que permite acelerar el acceso a tratamientos.

Big data tiene mucho que ofrecer al sector educativo. Los análisis pueden utilizarse para

seguir el progreso de los estudiantes e identificar estudiantes en riesgo. Los datos también son útiles para mejorar sistemas de evaluación y soporte de personal. Es difícil decidir entre todos los currículos, aunque el feedback de big data lo facilita mucho.

Algunas universidades ya usan sistemas de gestión de aprendizaje para hacer el seguimiento del progreso de los estudiantes. Registran el tiempo que dedican los estudiantes en diferentes páginas web y controlan el progreso mensual y anual.

Big data ofrece mayor precisión en los registros. Esto beneficia directamente al sector de la salud. Los registros más precisos son solo la punta del iceberg. Otro beneficio es el rápido acceso a historiales de pacientes, recetas y planes de tratamiento.

Imagine datos sobre salud pública instantáneos y gratuitos. Imagine tener suficiente información para reducir la propagación de enfermedades crónicas, o infecciones peligrosas de transmisión aérea...gracias a big data estos sueños son realidad.

La mejora de la eficiencia de la cadena de suministro es una ventaja obvia para el sector de la fabricación. Las estadísticas también muestran cómo big data mejora la calidad y cantidad de la producción, mientras reduce los residuos. La aceleración de la resolución de problemas y la mejora de la toma de decisiones también son resultado de big data en la fabricación.

Big data permite a los minoristas analizar tendencias en el comportamiento de compra examinando información de puntos de venta, tarjetas de fidelidad e identificación de radiofrecuencia. Lo que mejora las relaciones con los clientes, la toma de decisiones y métodos de transacción.

No hay duda de que big data ofrece muchos beneficios en muchas industrias. ¿Dónde puede usar su empresa esta tecnología para mejorar su rendimiento?

Big Data y la seguridad

[Título del tema: Big Data y la seguridad] Los ciberdelincuentes probablemente tienen decenas de formas de apropiarse de sus datos personales. El riesgo viene de software diseñado o cifrado de forma deficiente, empresas que no protegen datos personales y servicios de redes no seguras. Lo más vulnerable es la seguridad de la infraestructura y privacidad de los datos. Recuerde, una brecha en la seguridad de big data podría afectar a millones de personas. Por suerte, existen varias buenas prácticas para asegurar big data. Examinemos algunas de ellas.

Debe hacerse un seguimiento de quién accede a sus datos, y la mejor forma es con monitorización en tiempo real. El acceso no autorizado se marca al instante. Si hay ataques especializados más serios, necesita un nivel de inteligencia contra ciberataques para detectar y evitarlos.

Es probable que esté almacenando big data en la nube. Necesita revisar sus proveedores en la nube. ¿Qué métodos de protección usan? ¿Hay penalizaciones si no mantienen su seguridad? ¿Quién más está utilizando este proveedor, y qué tienen que decir?

Siempre es posible que experimente fugas de datos. Una buena práctica consiste en retirar siempre información personal de datos almacenados. Si bien hacer que los datos sean anónimos es un buen comienzo, el cifrado es crítico. Sus datos sin procesar y los datos de sus análisis deben cifrarse. El cifrado basado en atributos proporciona un control preciso del acceso a datos cifrados. Es un cifrado de clave pública, con la clave secreta del usuario y texto cifrado dependiendo de atributos, como por ejemplo dónde vive el usuario.

Otro método es realizar evaluaciones periódicas del riesgo. Si está recopilando datos privados, ¿qué políticas aplica para asegurarse de su seguridad? ¿Cómo se asegura de que su derecho a la privacidad no se incumpla? También podría compartir datos con otra empresa. Considere cómo se comparten exactamente los datos. Piense también sobre su seguridad. La reputación de ambas empresas está en juego, por lo que tiene que pensar en formas de reducir el riesgo. Cualquier dato publicado deliberadamente o accidentalmente que viole las políticas de privacidad del usuario puede derivar en responsabilidades legales. Dispositivos físicos, como su teléfono, son difíciles de asegurar. En cualquier momento que deje su teléfono en la mesa de un restaurante, o en el trabajo, alguien podría llevárselo o intentar acceder a él antes de que vuelva. Hay formas de superar la pantalla de bloqueo y contraseña. Por este motivo debe centrarse en la seguridad de la aplicación y no en la del dispositivo.

Además, puede considerar aislar los dispositivos y servidores que contienen datos críticos. La última, y quizás la más obvia medida de seguridad es usar software antivirus. Los paquetes antivirus suelen venir con algunas funciones de seguridad.

La protección de datos no es fácil. Vivimos en un mundo donde incluso datos despersonalizados pueden pasar por sistemas que los crucen con conjuntos de datos existentes, completando los identificadores. Por lo tanto, tiene que hacer lo necesario para mantener la seguridad de los datos.

Fases del ciclo de vida de los datos

[Título del tema: Fases del ciclo de vida de los datos] El ciclo de vida de una porción de datos puede ir de segundos a décadas. El tipo de datos de una porción de datos juega un rol importante; cada organización utiliza un ciclo de vida de datos diferente. Dicho esto, existe un ciclo de vida de big data universal. Las porciones de datos pasan por cinco fases básicas. En algún lugar, alguien, o algo, crea una unidad de datos. Podría tratarse de un sensor que registre la velocidad del viento, una entrada de datos con un clic en guardar en una nueva plantilla, o envío de mensaje de texto a un amigo. Al cargar un documento en un servidor, pasa a través del firewall y se copia o crea en la base de datos de su empresa. Estos son nuevos datos, nunca existieron antes. Incluyen cualquier dato que recopile la empresa.

Los nuevos datos creados se comparten. Esto se produce dentro de su empresa y con varias partes externas. Las redes sociales hacen que compartir datos sea una actividad pública.

Una vez la información esté en Internet, no se trata realmente de si se están compartiendo sus datos, sino de cuándo. El momento en el que sus datos abandonen un sistema interno de la empresa, no hay modo de corregirlos o recuperarlos.

Ya sea interna o externamente, alguien va a usar sus datos. Las empresas usan datos como información para sus procesos. Antes de cargar algo, tendría que haber aceptado, probablemente no leído, los términos del servicio o política de privacidad de la empresa. Esto es un acuerdo legal que explica cómo la organización manejará sus datos. La mayoría de las personas solo aceptan que sus datos se recopilen continuamente. Aunque esperan saber cómo se utilizan. Nadie desea que se venda su información a sus espaldas.

¿Qué ocurre cuando no se van a utilizar nunca más? La siguiente fase en el ciclo de vida de los datos es dónde se almacenan los datos en un archivo. Esto determina la vida de sus datos. Las normativas limitan el tiempo de archivo en algunos sectores, como en las industrias financiera o de la salud de EE.UU., aunque el almacenamiento en la nube

significa que los datos se mantendrán para siempre, lo que no es caro de hacer. Eso implica que las organizaciones pueden hacer análisis a largo plazo.

Aunque los datos también deben morir. Puede destruir datos borrando copias individuales de una base de datos. Esto se llama purga. En función de los términos del servicio, podría solicitar que se destruyan sus datos, aunque no siempre. Las organizaciones suelen mantener archivada alguna forma de sus datos.

Cada porción de datos individual pasa por un ciclo de vida de los datos. Puede durar momentos o décadas. Con más o menos espacio para estar disponibles, algunos datos podrían vivir para siempre en la nube.