

La empresa basada en datos

Por Mark Schwartz, estratega empresarial, AWS





Introducción

En la actualidad se habla mucho de la empresa basada en datos y la necesidad de convertirse en una. Pero, ¿qué significa exactamente convertirse en una empresa basada en datos y por qué es tan importante en el entorno digital actual? ¿Qué medidas prácticas puede tomar una empresa para incorporar los datos a su mentalidad y sus procedimientos? ¿Qué relación existe entre los datos y esa otra prioridad de la era digital: la agilidad empresarial y técnica? En este libro electrónico se describirá el concepto de la empresa basada en datos y se proporcionarán ejemplos de cómo las empresas utilizan los datos para impulsar sus negocios. También se establecerá la relación entre estar basado en datos y la agilidad, la transformación digital y la innovación continua.

Las organizaciones basadas en datos buscan **fundamentar sus decisiones estratégicas de negocios** en la evidencia que proporcionan los datos, lo que requiere un cierto rigor y, al mismo tiempo, la capacidad de innovar mediante la identificación, dentro de los datos, de las oportunidades que puedan llevar a nuevos productos o mercados. También tratan los datos como un recurso que pueden utilizar para mejorar las interacciones con el cliente y aumentar la eficacia. Es decir, **analizan** los datos para fundamentar la toma de decisiones y los **utilizan** para servir a sus clientes. Por ejemplo, los datos pueden ser la base de la personalización, la dinamización de los precios, la expansión de los mercados, la innovación de los productos o la optimización de la cadena de suministro.

Pero hasta hace poco, a las empresas les resultaba difícil utilizar los datos de estas formas, porque solo los consideraban en el contexto de las transacciones; como resultado, los bloqueaban en bases de datos aisladas que resultaban perfectas para el procesamiento de transacciones pero no tanto para el análisis ilimitado. Nuestro modelo mental era el de una factura o de un pedido: "Necesito 20 widgets a un precio de 100 USD por widget". O bien, "Abone 20 widgets a razón de 100 USD por widget". Los datos eran activos e imperativos, un estímulo u objeto para realizar una transacción. En la actualidad, el valor de los datos va mucho más allá de su papel transaccional.

¿Cómo podemos pensar en este valor en términos financieros y cómo podemos maximizarlo?

El valor comercial de los datos

Cualquier dato se puede utilizar en diferentes análisis para impulsar los resultados empresariales. Por tanto, tiene valor a la hora de materializar los resultados que se obtienen de esos análisis. Por ejemplo, si la empresa analiza el histórico de sus transacciones y como resultado, encuentra formas de optimizar su cadena de suministro, los datos han contribuido a permitir esa reducción del coste. Esto quiere decir que los datos tienen un valor empresarial derivado de su posible utilización para aumentar los beneficios o lograr los objetivos.

No es difícil encontrar casos de datos que se utilizan por su valor no transaccional. En Johnson & Johnson, por ejemplo, se utilizan los datos transaccionales almacenados en la nube para mejorar el cumplimiento de las buenas prácticas por parte de los médicos, optimizar la cadena de suministro y descubrir nuevos fármacos. Nike recopila datos acerca de los logros de los clientes para impulsar la experiencia digital del cliente en NikePlus. Lyft recopila y almacena las coordenadas GPS de todos sus viajes; cuando las analizaron, comprobaron que el 90 % de los viajes se solapaba con otros viajes desde ubicaciones cercanas. Esta información propició la creación de Lyft Line, un servicio que permite a los pasajeros compartir un coche y recibir descuentos de hasta el 50 %.¹

Como estos usos pueden generar ganancias en el futuro, aunque aún se desconozcan, los datos pueden ser considerados un recurso financiero (aunque, en la mayoría de los casos, no de tipo PCGA) Por tanto, no sorprende que los datos acumulados por una empresa puedan ser un factor de su valor de adquisición, o bien permitirle asociarse con otros proyectos empresariales. Considere, por ejemplo, la adquisición de LinkedIn por parte de Microsoft, con datos de 433 millones de clientes, por 26 200 millones de USD, o los procedimientos concursales de Caesars Entertainment Operating Corp. Inc. entre los años 2015 y 2017, cuando los acreedores argumentaron que los datos sobre los 45 millones de clientes afiliados al programa de lealtad, Total Rewards, representaban un valor de 1000 millones de USD y eran su recurso más valioso.²

Resulta útil pensar que el valor empresarial de los datos es similar al de una opción de compra financiera. Esto significa que los datos permiten realizar cambios en la cadena de suministro o lanzar un producto nuevo, pero no implican la obligación de hacerlo. Esa opción se puede ejercer o no, en función del valor que tenga la nueva empresa de acuerdo con los datos. Y aquí es donde surge el problema de identificar el valor del recurso de los datos: valorar una opción de compra es mucho más complicado que calcular el RSI de una corriente proyectada de flujos de efectivo. Como resultado, las empresas suelen desatender el valor; pero como demuestro en mi libro *War and Peace and IT (La guerra y la paz y la informática)*³, muchas de las técnicas de informática ágil generan este tipo de valor de opción.

¹ Casos prácticos de AWS. Consulte <https://www.youtube.com/watch?v=6A1tOFqvgek>, <https://aws.amazon.com/products/databases/> y <https://aws.amazon.com/solutions/case-studies/lyft/>

² Ambos ejemplos de <https://sloanreview.mit.edu/article/whats-your-data-worth/> Puede encontrar un análisis detallado de la quiebra de Caesars en <https://turn-around.org/sites/default/files/11.%20Paper%20-Caesars.pdf> La quiebra fue extremadamente compleja y el valor de Total Rewards se incluyó con otros activos, por lo que no está claro qué valor se le asignó en última instancia.

³ Mark Schwartz, *War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age (La guerra y la paz y la informática: liderazgo empresarial, tecnología y éxito en la era digital)* (Portland, OR: IT Revolution Press, 2019).



Datos y agilidad

El valor se crea no solo por los datos en sí mismos, sino también por las herramientas y los procesos implementados para analizarlos y generar esos resultados empresariales. En el mundo digital actual, plagado de cambios rápidos, incertidumbre y complejidad, de interrupciones podríamos decir, es necesario utilizar los datos como complemento de la agilidad empresarial y como respuesta rápida y flexible a las circunstancias cambiantes. La agilidad es lo que permite a las organizaciones convertir los cambios rápidos en oportunidades y evitar las interrupciones mediante la respuesta ágil a la amenaza de la competencia. En la era digital, las empresas han aprendido que es necesario contar rápidamente con versiones iniciales de los productos y hacer que evolucionen mediante los comentarios continuos del mercado.⁴

Los últimos años han aportado técnicas para incorporar agilidad al **proceso** de desarrollo de productos como, por ejemplo, el desarrollo de software Agile, DevOps y Lean. Se ha utilizado la nube para acelerar la entrega de funciones de TI, tanto para software como para hardware. Las estructuras organizativas basadas en equipos han posibilitado la movilización de los recursos para satisfacer las necesidades cambiantes. Todos estos desarrollos han permitido que las empresas aumenten la agilidad de sus procesos.

Pero los procesos ágiles son solo una parte de la historia: los propios datos de la empresa también deben ser ágiles. Deben estar disponibles con facilidad para usos inesperados y en constante cambio. Deben ser accesibles y significativos. Los empleados deben disponer con facilidad de herramientas para trabajar con los datos y de las aptitudes para hacerlo. Es la capacidad de hacer uso de los datos de manera flexible, para que estén disponibles para nuevos usos desconocidos hasta ahora. Es el eslabón perdido para lograr agilidad empresarial y permite diferenciar una organización ágil de otra que simplemente haya adoptado las estructuras y la parafernalia de los modelos ágiles. La agilidad empresarial exige que los datos sean ágiles. Una empresa basada en datos domina ambos aspectos.

El enfoque de dotar de agilidad a los datos es nuevo. Mientras los datos eran solo transaccionales, se podían aislar en bases de datos altamente estructuradas que reflejaban cómo se utilizarían para esas transacciones. Nuestras herramientas eran bases de datos relacionales como Oracle o SQL Server, cuyo punto fuerte es el procesamiento transaccional. Utilizamos los datos para realizar las propias transacciones y para generar informes operativos que las complementaran.

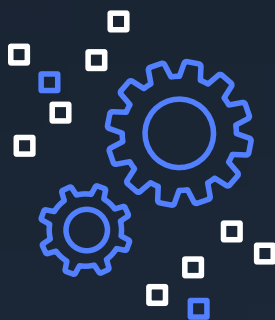
Cuando prestamos atención a la privacidad, la aplicamos limitando de manera estricta el acceso a los datos, en lugar de buscar formas en que los datos estuvieran disponibles dentro de los límites de protección de la privacidad. En lugar de "privacidad por diseño", practicamos una especie de "privacidad por falta de claridad".

Y sí, fueron varios los intentos de liberar datos para realizar análisis ad hoc mediante los denominados sistemas de inteligencia empresarial (BI). Pero las herramientas actuales han superado con creces el objetivo inicial de los sistemas de BI: ahora tenemos el aprendizaje automático, una variedad de bases de datos para fines específicos que controlan distintos tipos de datos, algoritmos para el procesamiento en paralelo masivo, grandes cantidades de datos no estructurados como vídeo y voz, dispositivos IoT que proporcionan flujos de datos derivados de sensores y, en definitiva, cantidades ingentes de datos. Con estas herramientas ahora podemos liberar los datos de su contexto transaccional y operativo.

En especial, nos hemos dado cuenta de que el estar basado en datos no es solo un desafío técnico sino también organizativo. **Para estar basada en datos, una organización debe cambiar su percepción acerca de cómo toma las decisiones empresariales y cómo interactúa con los clientes. Es un compromiso con el valor de los datos, una especie de humildad organizativa que reconoce que "los datos saben más que nosotros".**

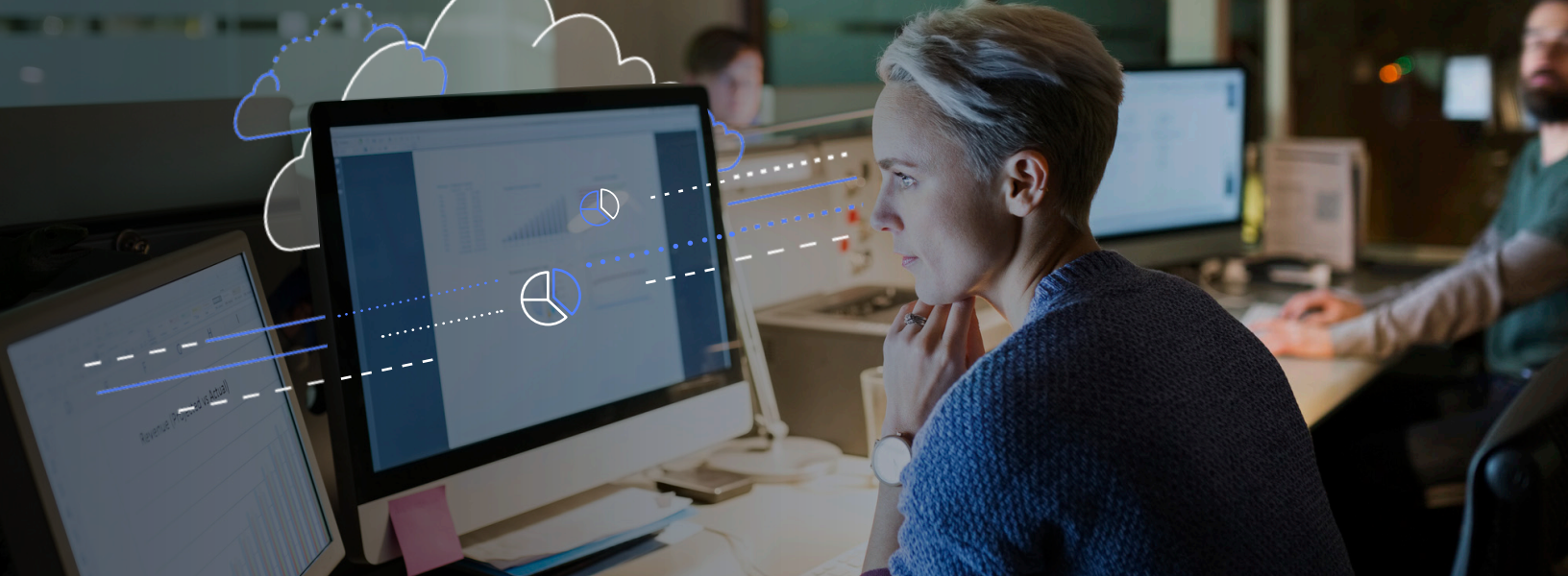
¿Qué se puede hacer para que los datos estén disponibles para ser utilizados de formas inesperadas, es decir, cómo se pueden utilizar para lograr la agilidad empresarial? ¿Cómo podemos aplicarlos para incorporar rigor y creatividad a la toma de decisiones empresariales? ¿Cómo se puede cambiar la cultura empresarial para aprovechar esta nueva flexibilidad?

¿Y cómo se pueden establecer controles adecuados para salvaguardar la privacidad de los datos al tiempo que se permite que sean utilizados de forma flexible y rápida?



En realidad son dos preguntas:

1. ¿Cómo podemos aportar agilidad a los datos?
2. ¿Cómo podemos utilizar los datos para dotar de agilidad a la empresa?



Agilidad para los datos

¿Cómo podemos aportar agilidad a los datos?

Para lograr la agilidad empresarial, es necesario estar dispuesto a responder a los cambios inesperados en los entornos empresariales y competitivos, así como crear innovaciones reales. Por este motivo, tendremos que poder trabajar con nuestros datos de formas que no siempre anticipamos cuando los obtenemos.

Nuestros desafíos:

- Es probable que nuestros datos estén almacenados de manera aislada en bases de datos relacionales y transaccionales, y que las distintas partes de la organización no puedan acceder a ellos.
- Es posible que no dispongamos de las herramientas de análisis adecuadas, o que estas no estén disponibles para las personas correctas en el momento preciso.
- Nuestros modelos de seguridad y privacidad son ad hoc, ya que seguramente nunca contemplamos el uso de los datos para su exploración. Muy probablemente, la privacidad se fomenta haciendo que los datos sean tan inaccesibles como sea posible.

Nuestras metas:

- Maximizar la disponibilidad de los datos, sujeta a la protección de privacidad y confidencialidad.
- Fomentar la transparencia en toda la empresa eliminando los conjuntos de información aislados.
- Ofrecer a los empleados las herramientas adecuadas para explorar los datos de formas no planificadas y que aprovechen los últimos avances en el campo del análisis.
- Asegurarnos de contar con la experiencia necesaria para interpretar los datos de forma rigurosa y creativa.

En “[Analytics without Limits: FINRA’s Scalable and Secure Big Data Architecture](#)”, John Brady, director de seguridad de la información de FINRA (Autoridad Reguladora de la Industria Financiera de EE. UU.), enmarca estos objetivos magníficamente al afirmar que quiere **reducir el coste de la curiosidad**. Se refiere al coste en el sentido más amplio del término, incluido el tiempo que se tarda en obtener conclusiones de los datos y el riesgo que supone hacer que estén disponibles. El cometido de FINRA es explorar los más de 37 mil millones de transacciones diarias que se producen en los mercados financieros, en busca de patrones de fraude. Como desconocen con antelación el aspecto de un patrón de fraude, dependen de la experiencia de sus analistas para detectar comportamientos sospechosos. Su labor se centra en la curiosidad: quieren que los analistas que examinan los datos sientan curiosidad por los patrones que aparecen y por qué. La labor de su organización de TI es reducir el coste de esa curiosidad y el esfuerzo que un analista debe invertir en explorar una corazonada.

La idea de Brady se aplica a todas las organizaciones y todos los roles. ¿Puede un vendedor explorar fácilmente los datos para buscar patrones inesperados en la actividad de compra de los clientes? ¿Pueden las operaciones explorar datos para identificar optimizaciones del rendimiento o diagnosticar problemas en los procesos operativos? ¿Pueden los departamentos financieros explorar los datos para idear nuevas formas de impulsar el rendimiento o diseccionarlos para impulsar la toma de decisiones ejecutivas? ¿Pueden los responsables de TI probar sus hipótesis sobre cómo optimizar el gasto en la nube con rigor y creatividad?

La curiosidad impulsa la innovación y la mejora. Los datos ágiles permiten a los empleados explorar ideas, presentimientos, hipótesis y conjeturas con libertad, así como promover nuevas ideas con datos que las complementen.

Para que los datos sean ágiles, una empresa debe considerar qué datos obtiene y el modo en que lo hace, cómo los conserva, cómo y en qué condiciones hace que estén disponibles, y con qué herramientas y aptitudes cuenta para trabajar con esos datos.



1

Obtener los datos



Para utilizar los datos con agilidad, primero hay que obtenerlos. Y dados los usos desconocidos a los que se verán sometidos, será necesario recopilar **más** datos de los que sabemos utilizar. En definitiva, ese es el concepto de "big data". Afortunadamente, con la nube, el coste de almacenar los datos es cada vez más bajo. Por tanto, podemos instrumentar los procesos empresariales de modo que se genere una gran cantidad de datos para su análisis. Por ejemplo, las aplicaciones del Internet de las cosas (IoT) incluyen con frecuencia sensores que emiten un flujo de puntos de datos a la nube y que las empresas pueden analizar de forma inmediata o almacenar para su análisis posterior. Ahora las empresas también pueden trabajar con una variedad de tipos de datos mucho mayor: vídeo, texto y voz, por ejemplo. Las posibilidades de utilizar toda esta información de formas interesantes y novedosas son enormes.

GE Oil and Gas, por ejemplo, introduce a sus oleoductos un dispositivo parecido a un escáner de resonancia magnética, denominado "cerdo", para recopilar más de 750 TB de información que facilita la detección de posibles problemas en su infraestructura. Hudl ha recopilado 10 PB de vídeo y otros datos que los entrenadores deportivos pueden analizar con los jugadores. En Peloton recopilan datos de sus ciclos de ejercicios y los analizan para proporcionar información detallada a sus clientes. Y Airbnb acumula aproximadamente 50 GB de datos diarios para el análisis rápido en la nube, y utilizan Amazon Elastic MapReduce (EMR), una herramienta que permite el análisis rápido y en paralelo de grandes volúmenes de datos. ⁵

2

Almacenar los datos



Una vez que se han obtenido los datos, es necesario almacenarlos para que estén disponibles para el análisis. Tradicionalmente, almacenábamos los datos en un formato estructurado en función de las expectativas sobre su uso transaccional. Por ejemplo, podríamos tener un campo en una base de datos para “cantidad del pedido” y otro para “precio unitario”. Podríamos recopilar datos para rellenar estos campos y guardarlos en los correspondientes espacios de la base de datos, con la certeza de que siempre podríamos multiplicar esos valores para obtener un precio total. Al forzar los datos en este tipo de molde, resultaron útiles para las transacciones, pero podríamos haber perdido información que podría haber sido útil para el análisis. Este era el modelo de base de datos relacional.

Las últimas décadas han estado dominadas por el uso de estas bases de datos relacionales, muy indicadas para procesar con eficacia antiguos volúmenes de datos transaccionales de formas ya conocidas (“multiplicar el precio unitario por la cantidad del pedido”). Pero cuando trabajamos con datos no transaccionales, a escalas de transacciones ingentes por Internet o con datos que no encajan con facilidad en “campos de datos” predefinidos, ahora existen alternativas mucho mejores y específicas diseñadas para la nube.

Por ejemplo, Amazon Timestream es una base de datos diseñada específicamente para administrar datos de series temporales (como los que genera un sensor industrial o los que se obtienen del seguimiento de la actividad del mercado en el tiempo); Amazon Quantum Ledger Database se centra en los datos que se utilizan en cadena de bloques (datos cuyo historial debe ser verificable mediante técnicas como la criptografía) y Amazon Neptune, diseñado para representar conexiones y relaciones complejas, como las redes sociales. Las empresas ya no están limitadas a lo que pueden encajar en un modelo relacional.

Mejor aún, para la agilidad, los datos que vayan a ser utilizados para su análisis por determinar pueden ser almacenados en un repositorio flexible denominado lago de datos, en el que cada fragmento de datos se almacena en el formato en el que ha sido recibido. La potencia del lago de datos radica en las herramientas que pueden ser utilizadas para analizarlo, que permiten combinar información heterogénea, mezclar datos estructurados y no estructurados, datos de distintos silos organizativos y datos en grandes cantidades. Las herramientas actuales pueden aplicar algoritmos de aprendizaje automático y análisis estadísticos, y trabajar con texto de lenguaje natural, vídeo y voz.

Es decir, el lago de datos cumple la necesidad empresarial de almacenar los datos antes de conocer todas las formas en que serán utilizados. En el lago se pueden incluir datos de distintos silos empresariales y analizarlos de forma conjunta. Podemos establecer una forma rápida de incorporar al lago datos de una empresa recién adquirida y, así, obtener transparencia en sus operaciones e integrar sus datos con los propios. La magia que hace todo esto posible es: (1) el bajo coste del almacenamiento, (2) la disponibilidad de herramientas que funcionan con datos heterogéneos y de estructura flexible y (3) la disponibilidad de servicios que insertan datos en el lago de datos con un ancho de banda alto y de forma asíncrona (solo hay que enviar los datos al almacén de datos en cuanto se reciben, donde llegarán lo más rápido posible, sin esperas, como un correo electrónico).

3

Hacer que estén disponibles



El siguiente paso para aportar agilidad a los datos consiste en hacer que estén disponibles, cuando y donde sean útiles. (Observe que no he mencionado cuando y donde se *necesiten*. Aquí hablo de agilidad e innovación). El modelo actual más habitual es del aprovisionamiento de autoservicio. Cuando un analista siente curiosidad, puede combinar una serie de herramientas y un subconjunto de datos que se van a analizar sin necesidad de solicitarlos y esperar a que alguien los proporcione. La libertad resultante permite al analista seguir una línea de pensamiento, un “flujo”, en lugar de un proceso alterno de arranques y paradas que destruye la creatividad o que, podríamos decir, aumenta el coste de la curiosidad. La nube facilita esto al permitir que nuevos entornos de trabajo se aprovisionen, se utilicen y, posteriormente descarten cuando ya no sean necesarios. También facilita la implementación de protecciones para la privacidad (como se verá más adelante).

4

Proporcionar herramientas



Una empresa basada en los datos pone a disposición de sus empleados las herramientas analíticas adecuadas con facilidad y rapidez, con frecuencia mediante un modelo de aprovisionamiento automático, como se ha descrito antes. Existe una amplia variedad de software y servicios disponibles: si quiere realizar consultas de estructura tradicional sobre los datos, por ejemplo, puede configurar un almacén de datos en función de los datos del lago de datos, o bien aprovisionar una herramienta que permita realizar consultas de tipo SQL directamente sobre el lago de datos.

Pero, en la actualidad, existen muchas más posibilidades. Por ejemplo, puede visualizar los datos con herramientas de modelado, construir escenarios y evaluar sus consecuencias. La revolución analítica actual se centra en la inteligencia artificial y el aprendizaje automático, lo que abre nuevas posibilidades para lo que podemos hacer con nuestros datos: predecir resultados, identificar anomalías, categorizar datos, analizar opiniones, detectar patrones, guiar robots y mucho más.

Por ejemplo, Capital One utiliza el aprendizaje automático para detectar fraudes al tiempo que mantiene altos niveles de servicio al cliente. T-Mobile utiliza el aprendizaje automático para mejorar su servicio al cliente lo que le permite predecir los artículos que serán más útiles para el cliente, y ponerlos rápidamente a disposición de los agentes del servicio al cliente. Sky News, durante su cobertura de la boda real británica, utilizó el aprendizaje automático de AWS para reconocer las caras de los famosos entre los asistentes e identificarlos para los espectadores. Y en la Fórmula 1, Major League Baseball y National Football League utilizan el aprendizaje automático para mejorar la experiencia del espectador de sus deportes.⁶

Para aplicar el aprendizaje automático, se entrena un modelo basado en conjuntos de datos anteriores y posteriormente se aplica a datos nuevos mientras se observa. En AWS, existen tres enfoques generales de aprendizaje automático: (1) utilizar un modelo previamente entrenado como Amazon Rekognition, que ya ha sido entrenado para reconocer objetos en imágenes, o Amazon Lex, entrenado para comprender las intenciones expresadas en lenguaje natural, (2) entrenar y aplicar un modelo propio basado en cualquiera de los algoritmos habituales que se utilizan para el aprendizaje automático con Amazon SageMaker, o bien (3) utilizar algoritmos y enfoques de entrenamiento propios, si tiene empleados con esa formación, y trabajar directamente con infraestructura de Amazon optimizada para el aprendizaje automático.

Con herramientas como estas, las empresas pueden dar rienda suelta a la creatividad de sus empleados y buscar nuevas formas para utilizar los datos.

5

Mejorar la formación



El siguiente elemento importante de la extracción de valor de sus datos consiste en asegurarse de que cuente con empleados que tengan las habilidades correctas... y sentido de la curiosidad. Esto explica que en la actualidad la demanda de científicos de datos sea tan alta. Es evidente que existen muchas herramientas disponibles, incluso para personas con pocas habilidades o experiencia estadística. Pero, para sacar el máximo partido de los datos y hacerlo con rigor, es importante contar con personas que sepan realizar inferencias correctas de los datos.

Como ejemplo sencillo, los que tenemos menos experiencia estadística solemos depender en exceso de los promedios, incluso cuando el examen de una distribución completa de valores puede proporcionar información importante. Recuerdo un caso, cuando era director de TI de USCIS, en el que intentábamos reducir el tiempo necesario para procesar determinados tipos de aplicaciones. Creamos paneles para realizar el seguimiento del tiempo de procesamiento medio, pero cada cambio que probábamos parecía tener un impacto mínimo sobre la métrica. Habíamos pasado por alto que el pequeño número de aplicaciones que despertaban preocupaciones de seguridad nacional o fraude tardaban mucho más en procesarse, lo que sesgaba la media. No teníamos forma de controlar el tiempo que tomaban. Aunque nuestras mejoras se aplicaban a la gran mayoría de casos, debido al promedio altamente sesgado no podíamos realmente apreciar su impacto. Cuando detectamos el problema y, por ejemplo, comenzamos a supervisar el tiempo de finalización del percentil 85, pudimos identificar el impacto significativo de nuestros cambios en la gran mayoría de los casos. Teníamos los datos, las herramientas y el acceso, solo nos faltaba la habilidad para hacer las inferencias correctas.

Las decisiones basadas en datos también se pueden fundamentar incorrectamente cuando los datos se presentan de forma confusa (aunque sea por accidente). En su libro *The Visual Display of Quantitative Information*, Edward Tufte muestra cómo los datos se pueden oscurecer o distorsionar en función de cómo se presenten.⁷ De nuevo, una empresa que quiera utilizar sus datos de forma rigurosa tendrá que asegurarse de contar con las habilidades correctas de análisis y presentación, además de los datos.

6

Proporcionar protecciones



Antes de poder hacer que los datos estén disponibles para nuevos usos, como por ejemplo para satisfacer la curiosidad, es necesario proteger su privacidad y confidencialidad. Las empresas basadas en datos practican la "privacidad por diseño" y establecen de forma deliberada medidas de seguridad basadas en la planificación y la previsión. Durante el proceso, para ganar velocidad y flexibilidad, se aseguran de que saben lo que deben proteger y establecen formas automáticas de hacerlo. De hecho, el reciente Reglamento General de Protección de Datos (RGPD) de la Unión Europea requiere la privacidad por diseño.

La nube proporciona muchas herramientas para configurar controles de acceso automatizados y lo hace a nivel granular, lo que permite asignar a los empleados el acceso precisamente a los datos que necesitan. Existen formas de realizar el seguimiento del origen y la validez de los datos, de cifrarlos u ocultarlos, y de restringir el acceso por campo o por registro. Es decir, usted puede especificar a qué datos de los clientes tiene acceso un empleado y qué fragmentos de datos asociados con esos clientes puede ver. Amazon Macie utiliza incluso el aprendizaje automático para identificar qué datos del lago de datos son información de identificación personal (PII) y realizar el seguimiento de su uso. O bien, usted puede optar por administrar datos solo en un nivel agregado o con información enmascarada o anonimizada. La flexibilidad está allí; las empresas basadas en datos deben tomar decisiones responsables sobre la privacidad en función del tipo de datos que controlen.

Al utilizar las grandes cantidades de datos de los que dispone la empresa surgen otros muchos desafíos. Uno de ellos es conectar de forma precisa datos de diferentes sistemas de TI que pertenecen a un único individuo, en especial en países como EE. UU. que carecen de un sistema de identificación nacional único. Los datos pueden ser imprecisos, no solo por errores surgidos durante su entrada, sino también por las limitaciones de los sistemas de TI que los recopilan. Por ejemplo, existen sistemas de TI que solo permiten un apellido y un nombre, lo que supone imprecisiones para las personas que tienen más de dos nombres.⁸

De cualquier forma, el objetivo de una empresa basada en datos es hacer que los datos estén disponibles para impulsar la toma de decisiones rigurosas y precisas, y la innovación continua. Esto pasa por recopilar y almacenar datos para utilizarlos con flexibilidad más tarde, y hacer que estén disponibles, junto con las herramientas adecuadas para quienes vayan a utilizarlos, mediante la garantía de la privacidad y la confidencialidad por diseño, la promoción de habilidades para realizar inferencias válidas y la resolución de problemas de higiene de los datos que puedan conducir a decisiones fundadas incorrectamente. Esto es lo que significa aportar agilidad a los datos.

¿Cómo podemos utilizar los datos para aportar agilidad a la empresa?

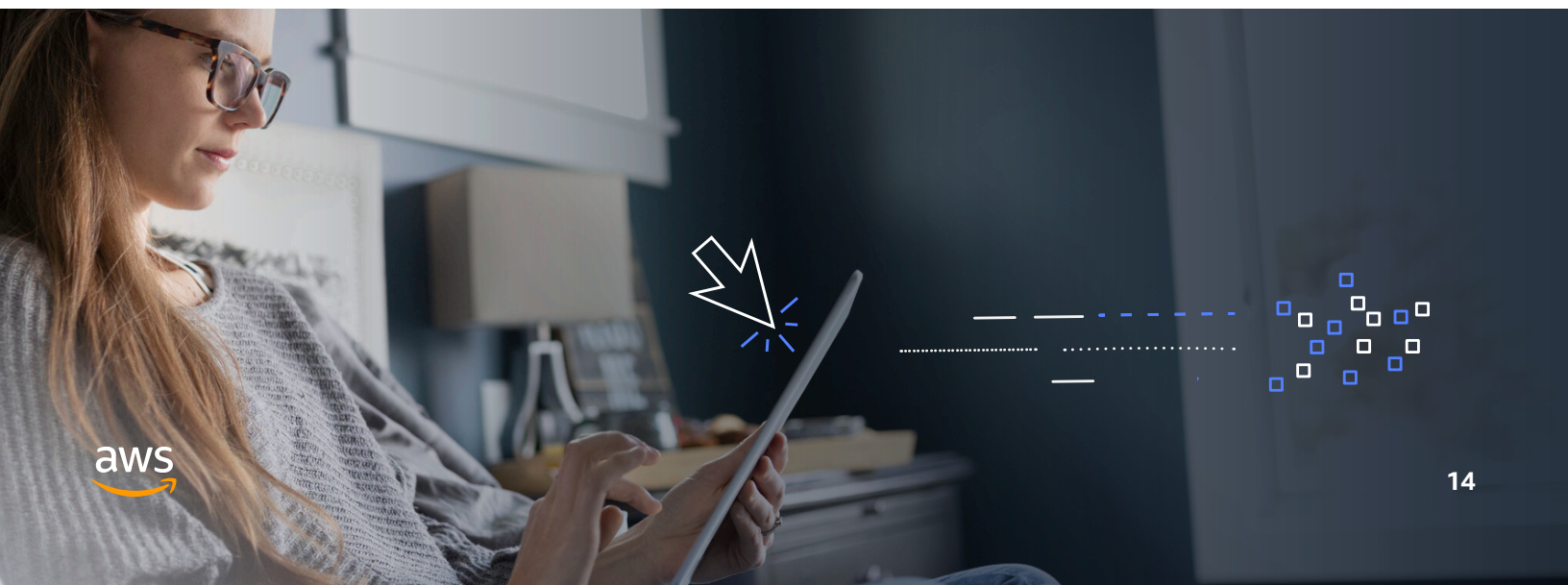
En la era digital, el funcionamiento de una empresa ágil consiste en probar una idea, obtener comentarios y ajustar el rumbo posteriormente, una y otra vez. Este enfoque de retroalimentación rápida permite que la empresa innove (con menor riesgo, mayor velocidad y a bajo coste) y reduzca el riesgo de la inversión probando las ideas antes de abrazarlas. El resultado es la combinación correcta entre los productos de la empresa y los mercados a los que atiende, y garantiza que la empresa resuelve el problema correcto de la forma adecuada en el momento preciso.

Retroalimentación rápida

En este sentido, los comentarios no significan preguntar a los clientes si les gusta una característica o un producto nuevo. Es habitual que las empresas basadas en datos utilicen comentarios cuantitativos, que se obtienen mediante la observación de cómo los clientes *actúan realmente*, o bien mediante la supervisión de cambios en el comportamiento del mercado u otras métricas.

Por ejemplo, las empresas suelen mejorar la facilidad de uso de sus sitios web mediante pruebas A/B, es decir, prueban dos variantes de un aspecto del diseño (normalmente una variante es la versión actual y la otra es un elemento de diseño nuevo que piensan introducir). Muestran a varios clientes la versión A y a otros la versión B. Recopilan datos acerca de la actividad de los clientes y los analizan en función de los resultados que les interesan. Para decidir si un botón tiene que ser de color verde o rojo para maximizar el número de veces que se pulsa, pueden mostrar la versión en verde a algunos usuarios, la de color rojo a otros y comprobar cuál genera más clics. Expedia y Netflix son ejemplos de empresas que realizan pruebas A/B periódicas y extraen grandes cantidades de datos de un almacén de datos en la nube.⁹

⁹ <https://www.youtube.com/watch?v=k8PTetgYzLA>.





El enfoque altamente eficaz de aprendizaje y adaptación a través de los comentarios va más allá de las pruebas A/B de la interfaz de usuario. Las ideas sobre productos nuevos, por ejemplo, se pueden probar mediante la creación de un “producto viable mínimo”, la versión más pequeña y sencilla del producto que la empresa puede utilizar para obtener información acerca de si tendrá éxito o no, o qué debe ser cambiado para que lo tenga. Las estrategias de marketing, promociones o alternativas tecnológicas se pueden probar mediante mediciones y pruebas para reducir la incertidumbre. Y la clave consiste en recopilar datos y hacer que estén disponibles para el análisis.

La técnica de utilizar productos viables mínimos y comentarios de retroalimentación rápida se describe en el libro *The Lean Startup* de Eric Ries.¹⁰ Según Ries, en un momento dado, una empresa emergente baraja dos hipótesis: una hipótesis de valor, sobre cómo su producto generará valor para los clientes, y una hipótesis de crecimiento, sobre cómo la empresa puede aumentar su mercado, es decir, conseguir que los clientes utilicen el producto. El producto viable mínimo es el producto más pequeño que proporcionará información a la empresa para confirmar o refutar estas hipótesis, para después poder realizar cambios y volver a probarlas en el mercado.

Este conjunto de procedimientos no se aplica solo a empresas emergentes o al desarrollo de productos nuevos. Se ha convertido en la forma esencial para las organizaciones, incluidas las grandes empresas, de lograr agilidad empresarial mediante el cambio de rumbo en función de sus descubrimientos. Si una empresa piensa en desarrollar un nuevo sistema de TI para que lo utilicen sus empleados, probablemente tenga una hipótesis sobre cómo ese sistema de TI proporcionará los resultados empresariales propuestos en su caso empresarial. Esa hipótesis se debe probar y se deben realizar cambios en función de lo que demuestren los datos.

Como resultado, la agilidad necesita datos; para aprender y adaptarse, la empresa debe recopilar datos sobre el impacto de sus nuevas iniciativas y utilizarlos para basar esas iniciativas. Además, la agilidad también requiere que la empresa detecte los cambios en su entorno para poder responder en consecuencia y maximizar sus resultados empresariales. Una empresa basada en datos no solo dota de agilidad a sus datos, sino que también los utiliza para complementar su agilidad.

Cultura y cambio de los procesos

En este sentido, estar basado en datos requiere otra forma de tomar decisiones, un profundo cambio cultural para muchas organizaciones. En el pasado, se tomaban decisiones mediante la elaboración de planes detallados, el análisis de opciones con los datos disponibles y la elección de la opción que, solo con los datos disponibles, parecía generar los mejores resultados. En el mundo digital, no aceptamos solo los datos disponibles en el momento de crear el plan. En su lugar, diseñamos experimentos para generar datos adicionales que son incorporados posteriormente a la toma de decisiones. La incertidumbre se resuelve mediante la generación de datos nuevos.

Un ejemplo es la técnica para la gobernanza de TI que diseñamos en USCIS. En lugar de redactar un extenso documento sobre los requisitos y pasarlo a los tecnólogos para su implementación, simplemente entregamos un objetivo empresarial. En una ocasión, por ejemplo, comprobamos que un procesador de casos (un "verificador de estado") podía procesar unos 70 casos al día y nuestro objetivo empresarial era aumentar mucho más ese número. En otro caso, comprobamos que se habían perdido varios archivos en papel al trasladarlos entre ubicaciones de procesamiento y quisimos eliminar esas pérdidas.

Para cada uno de estos objetivos, primero creamos un panel en el que se mostraba la métrica clave: el número de casos diarios o el número de archivos perdidos. En lugar de redactar un documento de requisitos, creamos un equipo multidisciplinar de operadores empresariales y tecnólogos de TI, y les encargamos mejorar la métrica. Les proporcionamos las herramientas para modificar con rapidez los sistemas de TI y los procesos empresariales, para después supervisar los paneles con ellos. Probaron con pequeños cambios incrementales y supervisaron diariamente los resultados. En función de lo que observaban en los datos, podían decidir el siguiente paso para maximizar el resultado. Y la dirección podía decidir si mantener la financiación de la iniciativa o dedicarla a algo distinto. El resultado fue un proceso de gobernanza sencillo, basado en datos y de riesgo reducido que proporcionaba valor rápidamente.

Esto lleva a otro aspecto importante: la responsabilidad se mejora mediante la transparencia. Al hacer que los datos estén disponibles con carácter general, el progreso del equipo se hizo visible. Como resultado, los organismos supervisores podían revisar de forma constante la decisión de inversión, para aumentarla o reducirla, redefinir los objetivos o cancelarla definitivamente. Los resultados eran el único medidor del éxito y se podían obtener con rapidez. Pero esos resultados tenían que complementarse con los datos.

DetECCIÓN DE PATRONES

Otro ámbito en el que los datos pueden promover la agilidad es la detección de cambios o el reconocimiento de patrones en el entorno. Por ejemplo, se puede utilizar el aprendizaje automático para detectar anomalías y responder ante ellas. Se puede entrenar un modelo de aprendizaje automático con datos históricos o rutinarios para que se acostumbre a lo que es "normal" y aplicarlo posteriormente para que busque actividades anómalas. Esta técnica puede ser utilizada, por ejemplo, para identificar transacciones fraudulentas o intrusiones de piratas informáticos en la red. O bien, para detectar equipos de la línea de producción de una fábrica con un comportamiento anómalo y que tengan que ser reparados o reemplazados, antes de que fallen.

Cuando se recopilan grandes cantidades de datos, es posible que se puedan identificar relaciones que no sabíamos que existían. Las empresas de redes sociales crean grandes bases de datos de relaciones entre usuarios. El Departamento de Seguridad Nacional podría descubrir que un posible terrorista al que investigan vivió en el pasado en la misma dirección de un terrorista ya identificado, lo que podría llevarles a formular preguntas cuando se topen con esa persona. Una serie de solicitudes de inmigración fraudulentas podrían proceder del mismo abogado. En este caso, se ha superado con creces el uso de los datos para el procesamiento de transacciones: ahora se pueden buscar relaciones interesantes e importantes entre esas transacciones. Pero, de nuevo, no sabemos qué relaciones exactas vayamos a encontrar: la agilidad, la flexibilidad y la curiosidad son las claves para obtener valor de los datos.

Para citar otro ejemplo de uso de los datos para "estar al tanto de los eventos", la existencia de un punto de datos puede confirmar que se ha realizado una actividad, por ejemplo, cuando se crean registros de auditoría de forma automática. Mediante el seguimiento de la pista de actividades, los auditores pueden validar el cumplimiento o investigar actividades inadecuadas. La cadena de bloques se suele utilizar para almacenar datos que confirman que se han producido actividades, por ejemplo una transferencia de dinero entre dos partes o la aprobación de un contrato por las partes interesadas. Mediante el uso de protecciones integradas y datos de auditoría para establecer el cumplimiento, las empresas pueden evitar los pesados procesos de cumplimiento que reducen la agilidad.

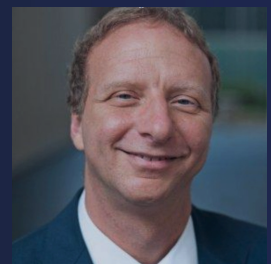
Evidentemente, el uso de los datos para respaldar la agilidad empresarial supone desafíos. Como se ha mencionado antes, hace falta preparación para obtener las inferencias adecuadas de los datos. Los datos no siempre indican qué acción se debe realizar: es necesario interpretarlos y tomar las decisiones correctas. A menudo, nos enfrentamos a un elemento de compensación entre falsos positivos y falsos negativos; por ejemplo, si utilizamos los datos a fin de detectar transacciones anómalas para identificar posibles fraudes, corremos el riesgo de marcar demasiadas transacciones como anómalas y molestar a los clientes, o no marcar las suficientes y no evitar el fraude. Cuanto mayor sea el conjunto de datos, mayor será la probabilidad de que surjan patrones sin sentido o de que los importantes queden ocultos entre la multitud de posibles conexiones. El ruido se acumula junto a la señal.

Conclusión

Una organización basada en datos es la que los utiliza para mejorar los resultados empresariales, tanto para impulsar un proceso de decisiones rigurosas como para que estén disponibles a fin de estimular la innovación y proporcionar valor a los clientes. Cuando los datos se limitan a un marco inflexible, se aíslan o son de difícil acceso, se crea una barrera para la agilidad empresarial, lo que impide que la empresa responda a las oportunidades o que comercialice sus productos con rapidez. Peor aún, cuando una empresa *no* controla sus procesos e inversiones a través del uso de los datos, resulta fundamental contactar con el mercado al que pretende servir o pasar comentarios que puedan ayudarle a mejorar el éxito de sus iniciativas. Una organización basada en datos, por su parte, utiliza los datos para lograr agilidad, y utiliza la agilidad para aumentar el valor de sus datos.

Acerca del autor

Mark Schwartz es estratega empresarial de Amazon Web Services y autor de *The Art of Business Value, A Seat at the Table: IT Leadership in the Age of Agility*, y de *War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age*. Antes de unirse a AWS, fue director de TI del Servicio de Ciudadanía e Inmigración de EE. UU. (parte del Departamento de Seguridad Nacional), director de TI de Intrax y director ejecutivo de Auctiva. Tiene un MBA de la universidad de Wharton, una licenciatura en ciencias informáticas de Yale y una maestría en filosofía de Yale.



Mark Schwartz,
estratega empresarial
de AWS

[Más información acerca de Mark Schwartz »](#)