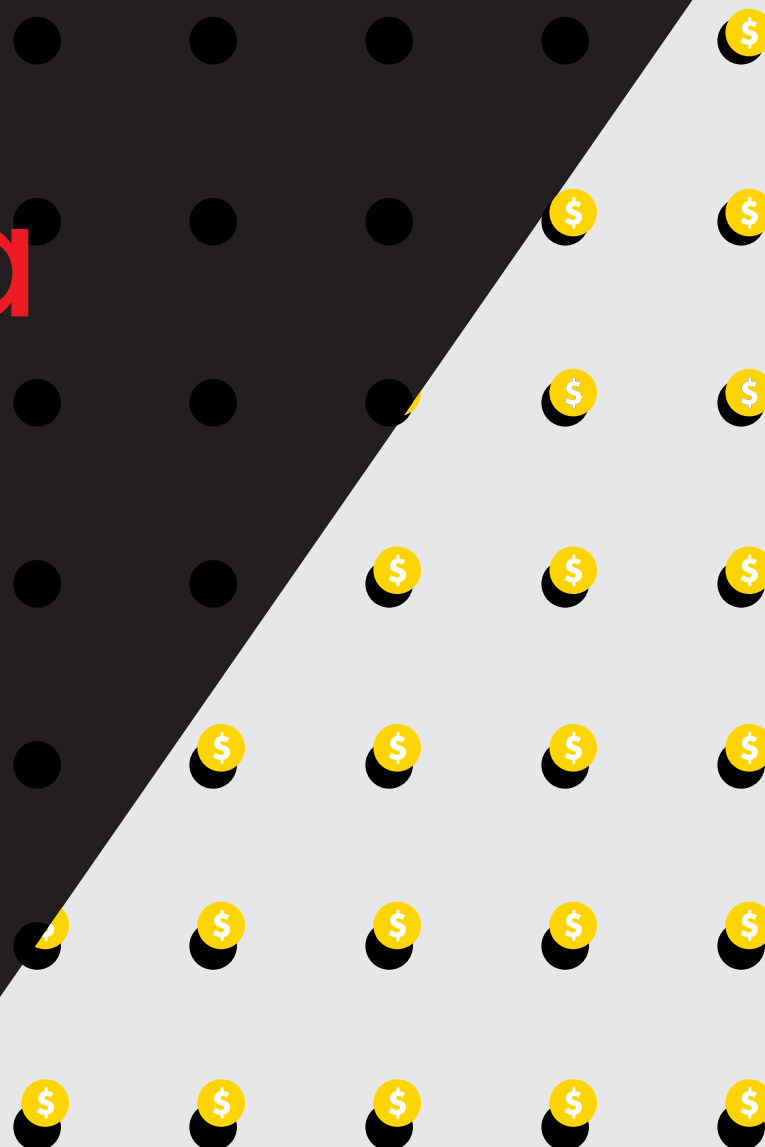


La importancia de los datos en la sombra

Tiene un filón a sus
pies... pero ni
siquiera lo sabe.



Tras la oscuridad, siempre brilla la luz

El mito de que apenas utilizamos el 10% de nuestro cerebro sigue vigente. A pesar de que innumerables neurocientíficos llevan años señalando que carece de fundamento, somos incapaces de desechar la idea de que, en nuestro interior, queda muchísimo por aprovechar.

Entonces, ¿qué ocurre cuando nos damos cuenta de que la mayoría de las empresas solo emplea un reducido porcentaje de su inteligencia?

Cuando, hace apenas unos años, explotó el mundo de los datos, las organizaciones se lanzaron a la carrera para sacar el máximo partido de sus nuevas y flamantes fuentes de datos.

Sin embargo, olvidaron que ya habían estado acumulando (y a un elevado precio) toda clase de valiosos datos sobre los procesos, los empleados, los clientes y los productos.

Todos esos datos residen en silos y almacenes heredados ubicados por toda la empresa. Como consecuencia, la mayoría de las empresas solo utiliza una pequeña parte de los datos cuya recopilación y cuyo almacenamiento ya ha pagado.

Ese es el filón por explotar.

Esa es la oportunidad que brindan los datos en la sombra.

No se trata solo de hallar información que podría cambiar las reglas del juego, sino también de descartar para siempre la increíblemente ineficiente y costosa práctica de acaparar datos a lo largo y ancho de la empresa.

Se trata, sin más, de arrojar algo de luz.

25-09-2014, 17:32:24

☀ ↑ 06:51	90°	Este	☾ ↑ 08:05	100°	Este
☀ ↓ 18:52	269°	Oeste	☾ ↓ 19:15	258°	Oeste

12 horas, 1 minuto (-3 min, 54 s)

Definición de los datos en la sombra

Son registros de máquinas, correos electrónicos, datos de aplicaciones, metadatos o datos de redes sociales.

Se trata del momento en que se adquieren los datos o se implantan las aplicaciones de las que proceden.

El hecho de que no tenga acceso a esos datos ni capacidad para perfilarlos significa que no tiene la menor idea de su grado de importancia o sensibilidad.

Los datos en la sombra son aquellos datos por los que la organización ya ha pagado, que ya ha recopilado y que ya ha almacenado en sistemas y almacenes diversos, pero que, en realidad, no se utilizan ni analizan, o ni siquiera se accede a ellos.

Es probable que los más antiguos sigan suponiendo un pozo sin fondo en su balance final porque nadie sabe qué datos contienen.

Definición de los datos en la sombra

Este imperdonable descuido:

- Le cuesta más (aún no sabe cuánto) de lo que debería.
- Favorece su exposición a una gestión ineficaz de los datos sensibles.
- Le impide obtener información exhaustiva sobre sus clientes, empleados, procesos y productos.

Una visión fiable de la empresa:

Si no puede utilizar los datos por los que paga, al final, carece de una visión coherente de la actividad de la empresa en su conjunto. Como consecuencia, no puede analizar los procesos, racionalizar los recursos ni localizar los datos que representan la ventaja propia de sus sistemas.

Uno de nuestros clientes, Interstate Batteries, una empresa de comercialización y distribución de pilas y baterías con un valor de 1.000 millones, tenía dificultades con la gestión y la previsión de inventarios. Por ello, la empresa no era capaz ni de elaborar previsiones para satisfacer la demanda del mercado de distribución ni de optimizar la toma de decisiones en cuanto a precios.

El suyo es un ejemplo más de que es imposible optimizar lo que no se puede analizar, e innovar si se ignora qué funciona.



Motivos de la falta de visibilidad de los datos en la sombra

Como es obvio, ninguna organización lleva a cabo actuaciones ineficientes, costosas o desacertadas a propósito. Aun así, a causa de la avalancha de aplicaciones y datos nuevos, demasiadas empresas pasan por alto los datos por los que ya han pagado y que ya les pertenecen.

University of Pittsburgh Medical Center (UPMC), por ejemplo, es una institución sanitaria mundial integrada que gestiona más de 1.200 aplicaciones. Eso implica caracterizar, integrar y analizar los datos ya existentes de tal modo que los investigadores tengan garantizado el acceso a todos los datos clínicos pertinentes sobre determinados pacientes.

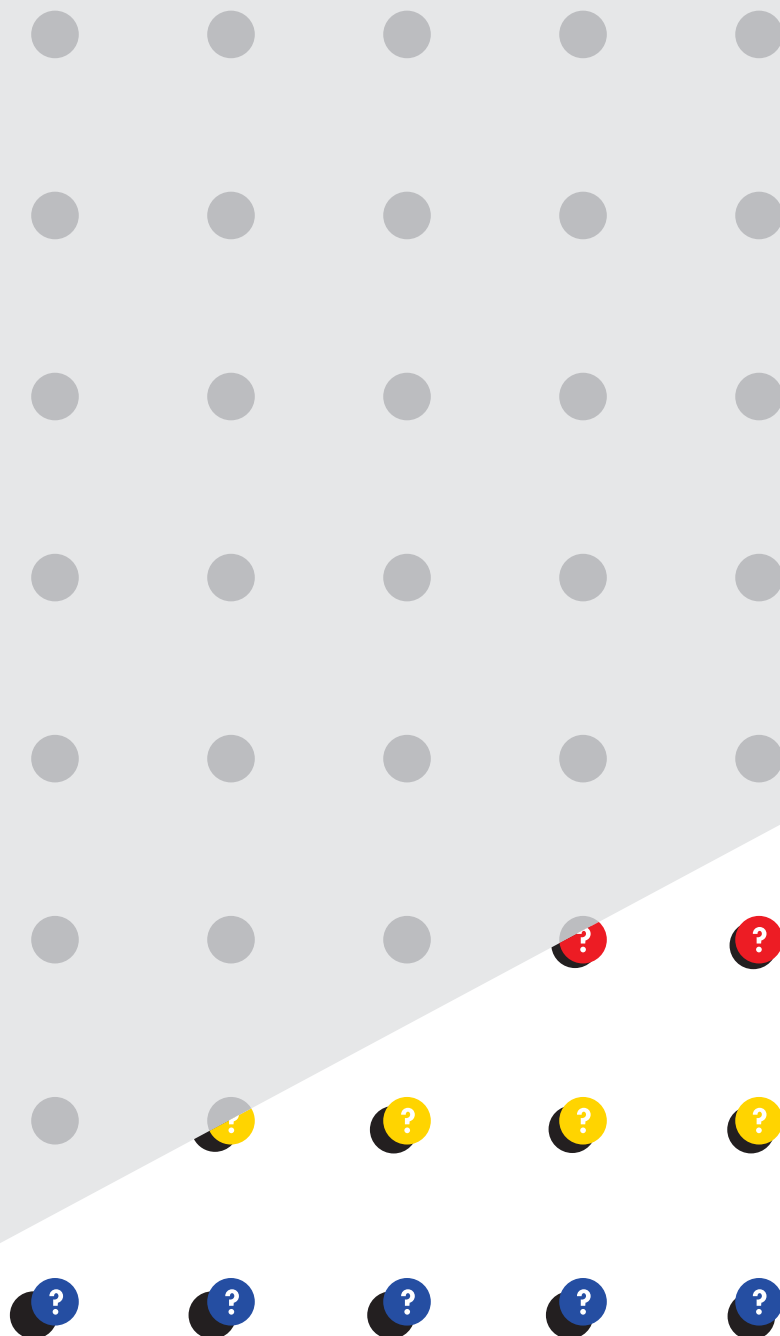
Sin acceso a dichos datos, UPMC sería incapaz de desempeñar su actividad. Veamos, a continuación, los cuatro motivos por los que las empresas pasan por tantas dificultades para utilizar los datos que tienen a su disposición.

1

Ni siquiera saben que los datos existen

Cuando gestores y equipos enteros tratan de resolver cuestiones difíciles o mejorar los métodos de trabajo, soslayan, demasiado a menudo, el reto que implica la búsqueda y el análisis de los conjuntos de datos con los que no están familiarizados. Por desgracia, es habitual que la falta de competencias, tiempo o ancho de banda les impida sacar a la luz los datos correctos.

Si los datos estuviesen almacenados de forma accesible para toda la organización, más equipos tendrían la capacidad de tomar decisiones mejor documentadas y poner a prueba un mayor número de hipótesis.

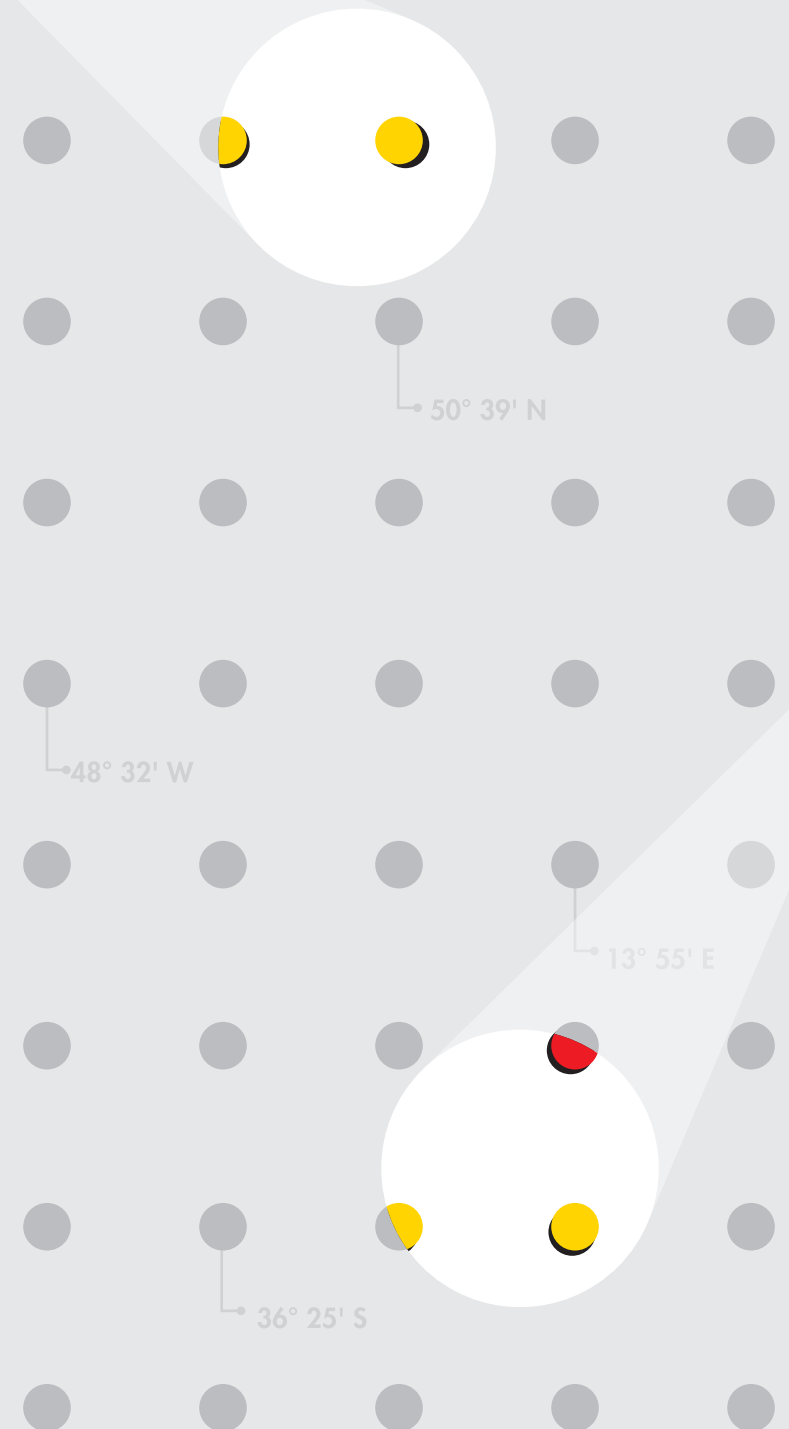


2

Saben que existen, pero ignoran dónde

El acceso a los datos entraña una especial dificultad cuando su propia arquitectura o bien la estructura organizativa se interponen en el camino. Si los departamentos trabajan aislados y los datos se guardan en almacenes heredados, hasta los equipos más curiosos chocan contra un muro impenetrable que está fuera de lugar.

Si no existe ninguna estrategia de almacenamiento y gestión de todos esos datos que abarque toda la empresa, la calidad de la toma de decisiones de la organización se verá mermada por las políticas internas y por tecnologías obsoletas.



3

Cuesta demasiado utilizar los datos

Aun cuando las organizaciones logran detectar los datos en la sombra necesarios, por lo general, tienen que hacer frente a una retahíla de costes relacionados con su procesamiento en los sistemas heredados (operaciones y unidades de procesamiento adicionales). Aunque ahorren en gastos replicando los datos en hardware más barato y usando estructuras de software novedosas como Hadoop, los costes iniciales asociados a la migración de los procesos y la adquisición de nuevas competencias suelen ser demasiado elevados como para realizar la inversión para un único proyecto.

Para aprovechar al máximo los datos que ya posee, es indispensable implantar las bases esenciales de una arquitectura de datos más moderna. Si no, seguirá pagando un precio demasiado alto por datos cuyo análisis no se puede permitir.



4

Cabe cierta legítima inquietud por la posible falta de cumplimiento en torno a algunos datos

Uno de los motivos más importantes por los que ciertos datos no se ponen a disposición de las herramientas analíticas es el miedo a que se produzcan fallos en la seguridad o se incumplan normativas relevantes sobre privacidad (como HIPAA en el sector sanitario). Aunque se trate de preocupaciones legítimas, resulta difícil justificar el coste que supone almacenar estos datos si no se van a analizar.

Para que su organización encuentre algún sentido a los datos que posee, es preciso implantar herramientas y procesos definidos con claridad que garanticen la seguridad y, en algunos casos, el anonimato de dichos datos.





Arroje luz sobre los datos en la sombra

Así pues, conocemos el elevado coste de los datos en la sombra en cuanto a los gastos reales, así como la posible y difícil de cuantificar pérdida de oportunidades. Es evidente que las organizaciones deben, como mínimo, tener la capacidad de acceder a sus datos en la sombra. Para lograrlo, tiene que dominar los ocho pasos siguientes.

Arroje luz sobre los datos en la sombra

Identifique los beneficios para el negocio

Si su proyecto no aporta innovaciones, optimiza los procesos, reduce costes ni mejora la ventaja frente a la competencia, le costará, y con razón, recabar los apoyos y el presupuesto que necesita. Tiene que describir con claridad y exhaustividad los beneficios para el negocio que reporta el acceso a los datos en la sombra.

Búsquelos

La mejor manera de empezar la búsqueda consiste en vincularla a un proyecto concreto e identificar los datos precisos para alcanzar los objetivos. Retroceda desde ese punto y confeccione una lista de los sistemas de origen, las aplicaciones e incluso los departamentos que podrían incluir u ofrecer los conjuntos de datos pertinentes.

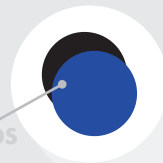
Extráigalos

La característica más destacable del Big Data es su diversidad. Por eso, no infravalore la diversidad de los datos en la sombra que va a intentar extraer: aprenda a recopilar datos con una transmisión en tiempo real con la misma eficiencia con la que los extrae de una base de datos. Además, los distintos almacenes de datos heredados presentan retos de extracción diferentes.

Identifique los beneficios para el negocio



Búsquelos



Extráigalos



Arroje luz sobre los datos en la sombra

Incorpórelos

Si introduce datos a raudales en los sistemas (en ocasiones, tras haberlos transformado), es imprescindible que documente las acciones realizadas y registre tantos metadatos como sean necesarios. En esta fase, también es importante que empiece a fijar (y mantener) estándares de coherencia semántica para los distintos formatos de datos.

Gobiérnelos

En vez de eludir la posibilidad de analizar los datos en la sombra, comprométase a cumplir las exigencias de las legislaciones oportunas. Defina la calidad de datos necesaria para que el proyecto funcione. Luego, establezca una política de control del acceso que se aplique mediante tecnologías de cifrado y enmascaramiento a fin de proteger los datos en origen. El objetivo consiste en gestionar los permisos otorgados a las distintas herramientas y los diferentes usuarios de negocio.

Almacénelos

La descentralización del almacenamiento de los datos es uno de los principales motivos por los que tantos datos permanecen en la sombra para la empresa. Piense a largo plazo y asegúrese de almacenar todos estos datos en un entorno gestionado lógico (centralizado o federado) y accesible. Esto tiene como finalidad evitar la duplicación y la proliferación innecesarias de datos en los sistemas. Así, contiene los gastos de almacenamiento y evita riesgos para la seguridad. En este sentido, la mejor apuesta es un data lake gestionado. (Lo ampliamos enseguida).



Incorpórelos



Gobiérnelos



Almacénelos



Arroje luz sobre los datos en la sombra

Prepárelos

Existen numerosos tipos de datos en la sombra dentro de todos los sistemas con distintos formatos y estructuras. Estos conjuntos de datos se deben cambiar de formato, analizar, filtrar, estandarizar, combinar y perfeccionar como preparación para su entrada en las aplicaciones y las herramientas analíticas.

Entréguelos

El paso final que debe dominar guarda relación con la facilidad para acceder a todos los datos y utilizarlos. Si se mantienen tanto el grado apropiado de coherencia semántica y la seguridad como los metadatos precisos sobre su linaje y sus características, los usuarios de negocio deben ser capaces de estudiar los datos con cualquier herramienta analítica. Según las aplicaciones a las que los incorpore y la naturaleza del proyecto, sabrá si los datos se deben entregar en tiempo real, en batch o en función de ciertos eventos.

En el caso de UPMC, hubo que coordinar esfuerzos e inversiones para que los investigadores pudieran integrar de forma electrónica (en primicia) los datos clínicos y del genoma de 140 pacientes que habían recibido tratamiento contra el cáncer de mama.

Prepárelos

Entréguelos

Proceso repetible para aprovechar los datos en la sombra

El mayor error que cometen casi todas las empresas consiste en creer que solo tendrán que recurrir a los datos en la sombra una única vez. No se trata de un proceso puntual. Los datos no hacen más que multiplicarse: en tamaño, en diversidad y en valor. Lo hacen de forma exponencial y continua. Además, el número y el tipo de aplicaciones a las que debe entregar los datos también varían de manera constante.

Por ello, en lugar de resolver proyectos individuales de datos en la sombra una y otra vez, conviene que se plantee crear un proceso repetible. Adopte la tecnología precisa para crear una infraestructura moderna que permita franquear el acceso a todos los datos y mantener su coherencia, de modo que disponga de datos limpios, seguros y conectados.

Ese es el motivo por el que tantas organizaciones líderes recurren a una solución de data lake gestionado.

STS-133
02.24.11

STS-101
05.19.00

STS-49
05.07.92

STS-65
07.08.94

STS-70
01.09.96

STS-73
11.22.89

STS-109
03.01.02

STS-108
12.05.01

STS-129
11.16.09

Proceso repetible para aprovechar los datos en la sombra

Un data lake gestionado es un lugar para gestionar el suministro y la demanda de los datos.

Gracias a él, los analistas de datos y los especialistas en datos cuentan con un entorno único de colaboración donde detectar, examinar, relacionar y adquirir cualquier tipo de dato de cualquier fuente interna de la empresa o ajena a ella; así, pueden preparar y disponer los datos para el análisis.

La finalidad es evitar la creación de una "ciénaga" en la que se limite a volcar todos los datos hallados en una ubicación sin tener en consideración la formación de un caos multiestructurado.

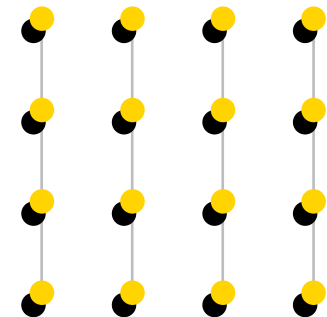
Un data lake gestionado se caracteriza por lo siguiente:

Suministra los datos con rapidez

Se garantiza la incorporación de todos los tipos de datos, su preparación para el análisis y su entrega a las herramientas de los usuarios de negocio de forma rápida y automatizada. En esencia, pone a disposición de los analistas todos los datos que necesitan.

Acata las directrices operativas de gobernanza de datos

Eso significa que sigue las normas de acceso y análisis de los datos que haya predefinido un órgano formado por personal de diferentes funciones y departamentos. Puesto que el data lake gestionado va a prestar sus servicios a toda la organización, debe tener representadas todas sus necesidades.



Proceso repetible para aprovechar los datos en la sombra

Conserva los metadatos descriptivos

En un data lake gestionado, los datos se controlan según el dominio y se almacenan con tantos metadatos contextuales como necesiten los usuarios de negocio antes de realizar sus consultas. A falta de este catálogo de contexto, los datos permanecerán bajo un halo de confusión.

Mantiene la calidad de los datos

Para que el data lake ofrezca un valor uniforme al negocio, es imprescindible confiar en la calidad de los datos. Por ello, debe mantener estándares de calidad de datos que estén a la altura de las expectativas de los usuarios de negocio y del uso previsto de los datos.

Registra el linaje de los datos

Para que los analistas no deban remontarse a las transformaciones efectuadas para crear el conjunto de datos a fin de comprenderlo, el data lake gestionado almacena la información sobre el linaje de los datos. De esta forma, es posible establecer enseguida la procedencia de los datos y maximizar su reutilización con la confianza de saber que los analistas no tienen que empezar de cero todo el rato.

STS-49
05.07.92

STS-33
11.22.89

STS-133
02.24.11

Proceso repetible para aprovechar los datos en la sombra

Protege los datos

La diferencia entre un data lake gestionado y uno no gestionado suele estribar en la capacidad de la tecnología para proteger los datos de forma adecuada. Se trata de controlar el acceso a los distintos datos mediante capas de seguridad como el control del acceso, el cifrado y el enmascaramiento de datos.

Mantiene la coherencia semántica

Para que los usuarios de negocio puedan llevar a cabo análisis de autoservicio de distintos conjuntos de datos, conviene garantizar que tales conjuntos estén preintegrados en el sistema. De ese modo, resultan fáciles de localizar e interpretar y no es necesaria una transformación constante.

Llegados a este punto, debe saber que la tecnología del data lake no ofrece las mismas funcionalidades que un data warehouse. No es compatible con la coherencia transaccional ni cuenta con ningún modelo específico para la elaboración optimizada de informes.

Dicho eso, el data lake aporta una visión de los datos en la sombra que no alcanza el data warehouse. Brinda la opción de examinar los datos antes de procesarlos en el data warehouse y los sistemas de business intelligence.

Se trata de una ventaja fundamental con respecto a la cruda realidad en que se encuentra la mayoría de las empresas, que llegan a tardar meses en suministrar datos a sus usuarios de negocio.

En la sección siguiente, veremos qué han hecho nuestros clientes para sacar el máximo partido de los datos ocultos en todos sus sistemas.



Casos prácticos sobre datos en la sombra

Si la organización tiene una visión fragmentada de la realidad del negocio, es imposible que mejore sus métodos de trabajo. Si tenemos en cuenta el montante gastado en almacenes de datos heredados como mainframes, bases de datos de aplicaciones, data warehouses y data marts, esto se convierte en una adivinanza de lo más costosa.

Por otro lado, si la organización no tuviera ningún dato oculto en silos y almacenes de datos inaccesibles, el personal podría formular más hipótesis y poner a prueba más corazonadas.

A continuación, presentamos parte de lo que son capaces de hacer nuestros clientes ahora que tienen acceso a la mayoría de los datos que poseen.

Serpens Caput

Ofiuco

Aquila

Casos prácticos sobre datos en la sombra

En el año 2010, el estado de Washington recibió una subvención federal destinada a mejorar el sistema estatal integrado de enseñanza. Tenía que reunir datos que residían en silos de diez departamentos distintos, entre otros, servicios sociales, recursos humanos y organismos educativos.

Cuatro años después, el estado ha reunido datos sobre el personal, la educación primaria y secundaria, y los estudios superiores con el objetivo de obtener una visión sencilla y longitudinal de los alumnos a lo largo del tiempo. Los legisladores, los investigadores, los docentes y los responsables de los programas han mejorado la calidad del sistema estatal de enseñanza. Ahora, las políticas y los programas se basan en un perfecto conocimiento de las variables por las que, al acabar sus estudios, algunos alumnos salen adelante y otros tienen más dificultades.



Casos prácticos sobre datos en la sombra

Ya hemos mencionado las dificultades que atravesaba Interstate Batteries: **gestionar y planificar inventarios con el fin de elaborar previsiones para satisfacer la demanda del mercado de distribución y optimizar la toma de decisiones en cuanto a precios.**

Ahora, la empresa toma decisiones ágiles sobre precios e inventarios y posee más capacidad de respuesta al sondear conjuntos de datos internos dispares y complementarlos con datos de fuentes externas no estructuradas como, por ejemplo, datos meteorológicos, precios de la competencia, tendencias de compra de automóviles y conversaciones de redes sociales.



Gestión de inventario en tiempo real

Casos prácticos sobre datos en la sombra

También le hemos contado la historia de cómo UPMC integró los datos clínicos y del genoma de 140 pacientes que habían recibido tratamiento contra el cáncer de mama. Vea el vídeo para saber cómo se emplea y estandariza la información y cómo se comparten los conjuntos de datos para ofrecer una imagen integral de los datos, lo cual facilita la posibilidad de tener una auténtica ventaja frente a la competencia y de comprender los efectos de la atención mucho antes de la posible aparición de la enfermedad.



Datos en la sombra y diagnóstico

Casos prácticos sobre datos en la sombra

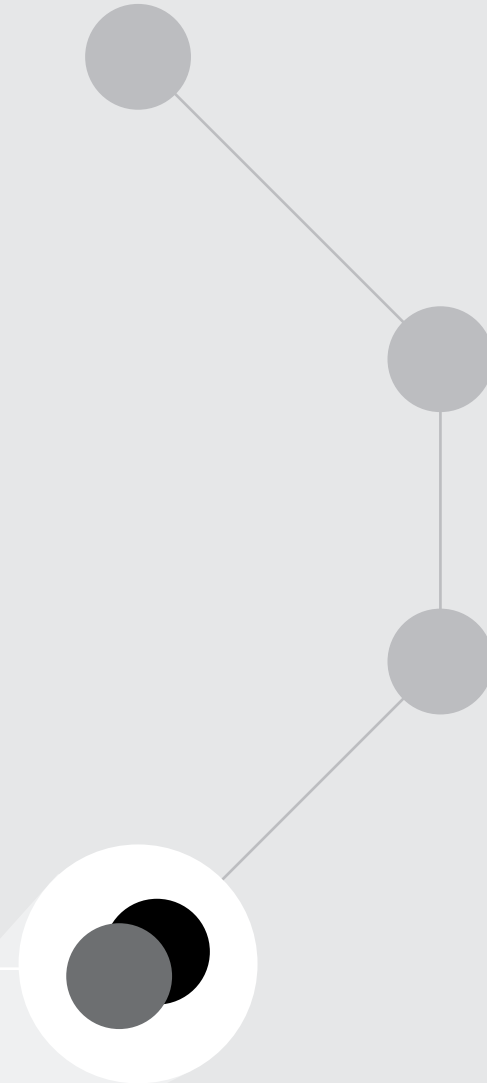
Una gran multinacional de servicios financieros ha creado un data lake gestionado con la intención de reducir costes y mejorar la eficacia de sus procedimientos para la prevención de fraudes y contra el blanqueo de capitales.

Una empresa del sector del gas y el petróleo ha creado un data lake para almacenar y procesar los datos procedentes de la actividad sísmica, los sensores, los dispositivos y los pozos. Así, la empresa puede integrarlos con los datos de las aplicaciones empresariales para reducir el coste de los servicios y del mantenimiento y, al mismo tiempo, aumentar la eficiencia de la producción de los pozos.

Las unidades de negocio de una gran empresa tecnológica trabajaban de forma independiente y jamás compartían datos entre sí. En la actualidad, sus especialistas en datos pueden sondear todos esos datos en un data lake para detectar patrones y tendencias que mejoren la experiencia de los clientes e identificar nuevos flujos de ingresos.

Cuando arroje luz sobre todos sus datos en la sombra, abrirá sus puertas a todo un mundo de posibilidades en cuanto a información y eficiencia.

¿Cómo será el suyo?



The background image shows an industrial facility at night. In the foreground, there are large, dark, cylindrical storage tanks with yellow safety railings. In the background, there are tall distillation columns and other industrial structures, all illuminated by various lights, creating a complex scene of industrial activity.

Ventaja de los datos en la sombra

Los datos en la sombra no solo son indicativos de la poca utilidad del gasto en tecnologías, sino también de los esfuerzos de la organización por utilizar los valiosos conocimientos acumulados.

Cuanto más tarde en modernizar su arquitectura de datos, más le costará gestionar la transición.

Ventaja de los datos en la sombra

Cuanto más tarde en modernizar su arquitectura y poner los datos a disposición de las unidades de negocio, más le costará facilitarles la información que necesitan.

Por un lado, es imposible saber cuándo afrontará el reto la competencia y se servirá de sus datos en la sombra para tomar ventaja y, por otro, está el problema del crecimiento exponencial e inexorable de los datos empresariales. Más del 90% de los datos mundiales se ha generado en los tres últimos años y no se aprecian trazas de que el ritmo vaya a descender a corto plazo.

Para que las organizaciones consigan, por fin, sacar réditos de todos los datos cuya recopilación llevan años pagando, dichos datos han de estar visibles, accesibles y listos para la consulta. Y deben ser datos limpios, seguros y conectados.

Una vez que consiga eso, podrá empezar a responder a las preguntas que los usuarios de negocio siempre han formulado. La diferencia es que ya los habrá dotado de los medios para basar sus decisiones en datos sólidos y dignos de confianza. Datos que resuelven cuestiones como las siguientes:

- ¿Cuáles son las mejores acciones que podemos llevar a cabo a continuación para reducir la rotación de los clientes más valiosos?

- ¿Cuál es el mejor indicador del fracaso de un producto?
- ¿Cómo reconozco de forma activa las causas habituales de las actuaciones de mantenimiento?
- ¿Qué actividades de posventa suelen inducir compras recurrentes?
- ¿Es posible identificar a los expertos internos en la materia por la cantidad de contenido que producen y almacenan en nuestros sistemas?
- ¿Nos indican las puntuaciones de las entrevistas qué empleados de RR. HH. son más hábiles para identificar a los candidatos más eficaces?

Otras lecturas

El camino para tener bajo control los datos en la sombra implica, por fuerza, la transición al Big Data. Y, si hay algo que sabemos sobre las transiciones al Big Data, es que son tan complejas como prometedoras.



Por eso, hemos compilado todas las recomendaciones sobre Big Data en un voluminoso cuaderno.

Consiga el gran cuaderno del Big Data.

Acerca de Informatica

Ayudamos a las empresas a gestionar sus datos para que obtengan de ellos un valor de negocio cuantificable. También estamos ayudando a algunas de las mayores empresas del mundo a examinar los errores más comunes en la gestión de datos para que tengan éxito en sus proyectos escalables y repetibles de Big Data.

Hablemos.

