



Cinco mitos sobre el data lake

A medida que las empresas son más conscientes del valor y la importancia de sus datos y de su capacidad para abrir la puerta a nuevas oportunidades y modelos de ingresos, se ven cada vez más inundadas de tecnologías que ofrecen el «mejor» enfoque para gestionar todos estos datos y obtener información.

Por lo tanto, existen motivos para que el data lake sea una de las arquitecturas de referencia que está despertando el interés de muchos. Sin embargo, también existen varios mitos sobre los data lakes que hacen que las empresas y los responsables de TI pierdan tiempo valioso investigando todas sus opciones.

1.

El data lake solo se puede implementar en un único cloud → Falso

El data lake no está limitado a una única ubicación o cloud. No se limita a la implementación local. Puede crear un data lake en varios clouds con opciones híbridas. Un data lake es una arquitectura de referencia que no depende de la tecnología. Se trata de un enfoque utilizado por las organizaciones para situar los datos en el centro de sus operaciones y, de este modo, facilitar el acceso a una variedad de tipos de datos a escala masiva y, a su vez, capacitar a los usuarios con análisis de autoservicio.

2.

Hadoop es el único data lake → Falso

Aunque, a menudo, el término «data lake» se asocia con Hadoop o con el almacenamiento de objetos orientado a Hadoop, la realidad es que un data lake podría desarrollarse y utilizarse eficazmente sin incorporar Hadoop. Por ejemplo, un data lake eficaz podría basarse en diferentes sistemas de gestión de bases de datos relacionales. Un data lake combina varias tecnologías para establecer sistemas de conocimiento que proporcionen una exploración de datos fluida para que los expertos en datos respondan mejor a las necesidades de las empresas.

3.

Puede utilizar los data lakes para volcar cualquier dato, sin necesidad de gobierno → Falso

Aunque el software y el hardware son componentes clave de una solución de data lake, son igualmente importantes la catalogación de datos, su calidad y los procesos de gobierno y gestión de datos.

Al igual que algunos almacenes de datos se han convertido en enormes agujeros negros de los que nunca salen grandes cantidades de datos, un data lake puede llegar a convertirse en un pantano de datos si no se implementan políticas de buen gobierno.

4.

El éxito de un data lake se mide por el acceso → Falso

Volcar datos en una ubicación central no es una solución analítica real. El objetivo es llevar a cabo análisis de datos que aporten conocimientos empresariales de peso, así como también descubrir nuevas fuentes de ingresos, modelos de retención de clientes o extensiones de productos.

Pero estos datos deben ser fiables, pertinentes y estar disponibles para todos los consumidores de datos. Un data lake necesita un catálogo de metadatos inteligente que pueda relacionarse con la terminología empresarial, trasladando los datos codificados crípticos y haciéndolos más comprensibles con el contexto. También se atribuirá al origen y la calidad de los datos de los activos de información —tanto estructurados como no estructurados— y al tejido de gobierno para, así, poder garantizar que la información esté protegida, estandarizada, gestionada eficazmente y sea fiable.

5.

El data lake es el sustituto del almacén de datos → Falso

El data lake puede incorporar varios almacenes de datos empresariales (EDW, por sus siglas en inglés), además de otros orígenes de datos, como los procedentes de las redes sociales o IoT. Todos ellos se reúnen en el data lake, donde se puede integrar el gobierno, lo que simplifica el descubrimiento de datos fiables a los usuarios de toda la organización.

Por lo tanto, un data lake aumenta los entornos EDW para permitir, facilitar o facultar a los expertos y analistas de datos para que exploren sus datos fácilmente y descubran nuevas perspectivas, conocimientos y oportunidades. Esto los ayudará a innovar e impulsar el crecimiento empresarial.



Las ventajas del data lake multicloud

No se deje llevar por la desinformación. Un data lake gobernado puede brindar acceso en toda la empresa a una amplia gama de datos estructurados y no estructurados, al tiempo que garantiza su fiabilidad y seguridad en cualquier lugar.

Cuando se optimiza para las necesidades de una empresa, un data lake gobernado puede agilizar los análisis y mejorar la precisión de los conocimientos porque:



Los datos se asientan sobre una base de infraestructura segura y fiable.



Los canales de información de datos controlados llenan el data lake con información fiable y, a continuación, documentan los activos de información, sus metadatos y el contexto empresarial. Este proceso ofrece un flujo de datos en tiempo real en el data lake.



La calidad, el origen y el linaje de los datos son bien conocidos.



Los datos se presentan en lenguaje empresarial, lo que permite a los expertos ponerse a trabajar inmediatamente, en lugar de perder tiempo tratando de descifrar los términos.



Los datos están debidamente clasificados, protegidos y gobernados.

La verdad está ahí fuera

La [cartera de gestión de datos de IBM](#) ha ayudado satisfactoriamente a los clientes a evitar las trampas y los mitos comunes sobre el data lake. Dicho de otro modo, los ha guiado a través de los pasos más cruciales para lograr una implementación exitosa del data lake gobernado.

Descubra las soluciones de data lake de IBM

Tome decisiones más inteligentes al aprovechar más tipos de datos de distintos orígenes.



Lea el libro electrónico

Conozca cómo los data lakes gobernados crean oportunidades para obtener información empresarial clave.

