



Cinco mitos sobre o data lake

À medida que as empresas se tornam mais conscientes do valor e da importância de seus dados e da capacidade que eles têm de abrir novas oportunidades e modelos de receita, elas vão sendo inundadas por tecnologias que oferecem a “melhor” abordagem para gerenciar todos esses dados e gerar insights.

Por um bom motivo, o data lake é uma dessas arquiteturas de referência que estão recebendo muita atenção. No entanto, há vários mitos sobre o data lake que fazem com que empresas e gerentes de TI percam um tempo precioso pesquisando opções.

1.

O data lake só é implementável em uma única nuvem → Falso

O data lake não é limitado a uma única localização ou nuvem. Não se limita à implementação no local. Você pode criar um data lake em várias nuvens com opções híbridas. Um data lake é uma arquitetura de referência independente da tecnologia. É uma abordagem que as organizações usam para colocar os dados no centro de sua operação, para assim facilitar o acesso a diversos tipos de dados, em enorme escala, e oferecer aos usuários uma análise democratizada dos dados.

2.

O Hadoop é o único data lake → Falso

Embora o termo data lake seja frequentemente associado ao Hadoop, ou armazenamento de objetos orientado ao Hadoop, um data lake pode ser desenvolvido e usado de forma eficaz sem incorporar o Hadoop. Por exemplo, um data lake eficaz pode ser baseado em diferentes sistemas de gerenciamento de banco de dados relacional. Um data lake reúne diversas tecnologias para estabelecer sistemas de insights, de forma a fornecer exploração ágil de dados para que os cientistas de dados atendam às necessidades dos negócios.

3.

Você pode usar data lakes para despejar qualquer dado, sem necessidade de controle → Falso

Embora o software e o hardware sejam componentes-chaves de uma solução de data lake, a catalogação de dados, a qualidade dos dados e os processos de controle e gerenciamento de dados são igualmente importantes.

Assim como alguns armazéns de dados se tornaram enormes buracos negros, dos quais grandes quantidades de dados nunca escapam, um data lake pode se tornar um “data swamp” - um pântano, em vez de um lago - se boas políticas de controle não forem aplicadas.

4.

O sucesso do data lake é medido pelo acesso que fornece → Falso

Despejar dados em um local central não é uma verdadeira solução de análise. A meta é executar análises de dados que produzam insights significativos sobre os negócios, assim como descobrir novos fluxos de receita, modelos de retenção de clientes ou extensões de produtos.

Porém, esses dados devem ser confiáveis, relevantes e disponíveis para todos os consumidores de dados. Um data lake precisa de um catálogo de metadados inteligente que possa se relacionar com a terminologia de negócios, movendo dados codificados incompreensíveis e tornando-os mais compreensíveis no contexto. Também atribuirá à fonte e à qualidade dos dados de ativos de informações estruturados e não estruturados, e malha de controle para garantir que as informações sejam protegidas, padronizadas, gerenciadas com eficiência e confiáveis.

5.

O data lake é um substituto para o data warehouse → Falso

O data lake pode incorporar vários data warehouses corporativos (EDW), além de outras fontes de dados, como redes sociais ou IoT. Todos eles se reúnem no data lake, onde o controle pode ser integrado, simplificando a descoberta confiável de dados para usuários de toda a organização.

Portanto, um data lake aumenta os ambientes de EDW para permitir, habilitar ou capacitar cientistas e analistas de dados a explorar facilmente seus dados e descobrir novas perspectivas, insights e oportunidades para acelerar a inovação e o crescimento dos negócios.



Os benefícios do data lake multinuvem

Não se deixe levar pela desinformação; um data lake governado pode fornecer acesso em toda a empresa a uma ampla variedade de dados estruturados e não estruturados, ao mesmo tempo em que ajuda a garantir que seja confiável e seguro em qualquer lugar.

Quando otimizado para as necessidades de uma empresa, um data lake governado pode acelerar a análise e melhorar a precisão dos insights porque:



Os dados ficam em uma base de infraestrutura segura e confiável.



Os feeds de dados controlados preenchem o data lake com informações confiáveis e, na sequência, documentam os ativos de informações, seus metadados e contexto de negócios, fornecendo fluxo de dados em tempo real para o data lake.



A qualidade, origem e linhagem dos dados são bem compreendidas.



Os dados são colocados em linguagem de negócios, permitindo que os cientistas de dados comecem a trabalhar imediatamente, em vez de penar com o significado de termos crípticos.



Os dados são devidamente classificados, protegidos e governados.

A verdade está logo ali

O [portfólio de gerenciamento de dados IBM](#) ajudou com sucesso os clientes a evitar armadilhas e mitos comuns sobre data lake, auxiliando-os a navegar por etapas críticas e garantir sucesso na implementação de data lake governado.

Descubra as soluções de data lake da IBM

Conduza decisões mais inteligentes capitalizando mais tipos de dados de mais fontes.



Leia o e-book

Explore como os data lakes governados criam oportunidades para obter insights fundamentais de negócios.

