

- Data lakehouse é um termo relativamente novo na indústria de tecnologia da informação, que combina aspectos de data lakes e data warehouses

- A ideia central de um data lakehouse é unir a flexibilidade e a capacidade de armazenamento de grandes volumes de dados de diferentes tipos, características dos data lakes, com a capacidade de gerenciamento, governança e otimização de desempenho para operações analíticas complexas, típicas de data warehouses

[Para conhecer mais diferenças entre data lake, data warehouse e data lakehouse, leia, aqui no 'Notícias do Seguro', a matéria "Navegando pelas águas da gestão de dados"](#)

Como funciona o data lakehouse

O data lakehouse é construído sobre uma arquitetura de armazenamento de dados que permite armazenar dados em seu formato bruto (como em um data lake) e, ao mesmo tempo, oferece mecanismos para gerenciar, acessar e analisar esses dados de forma eficiente (como em um data warehouse).

Essa capacidade é alcançada por meio da implementação de camadas de gerenciamento de dados que incluem esquemas, índices, otimização de consulta e controle de acesso, diretamente sobre o repositório de dados bruto.

Benefícios do data lakehouse

1. **Flexibilidade e escalabilidade:** o data lakehouse permite o armazenamento de grandes volumes de dados de diferentes tipos e formatos, proporcionando escalabilidade e flexibilidade.
2. **Custo-eficiência:** utilizando a infraestrutura de armazenamento de baixo custo dos data lakes, o data lakehouse reduz os custos associados ao armazenamento de grandes quantidades de dados.
3. **Gerenciamento e governança:** o data lakehouse oferece recursos de gerenciamento e governança de dados avançados, melhorando a qualidade e a segurança dos dados.
4. **Desempenho analítico:** o data lakehouse otimiza o desempenho de operações analíticas e de consulta, mesmo em grandes conjuntos de dados, graças à implementação de recursos típicos de data warehouses.

Riscos ao não usar um data lakehouse

1. **Ineficiência em operações analíticas:** sem um data lakehouse, as organizações podem enfrentar dificuldades em realizar análises complexas de forma eficiente devido à falta de otimizações de consulta e à estrutura inadequada de dados.
2. **Problemas de governança de dados:** a ausência de um sistema que ofereça boas práticas de governança pode resultar em problemas de qualidade, segurança e conformidade dos dados.
3. **Custos maiores:** sem as otimizações e a gestão eficiente de dados oferecidos por um data lakehouse, as organizações podem ter custos operacionais e de armazenamento mais elevados.
4. **Perda de vantagem competitiva:** a incapacidade de analisar rapidamente grandes volumes de dados de diferentes tipos pode resultar em uma perda de insight e, conseqüentemente, de vantagem competitiva no mercado.

À medida que os volumes de dados continuam crescendo e a análise de dados se torna cada vez mais crucial para as organizações, o data lakehouse surge como uma estratégia importante para o gerenciamento eficaz de dados

Fonte: CNseg, em 20.05.2024