يُشبه كلاهما مستودعات بيانات ضخمة، لكنهما يختلفان في الهيكل والغرض. يخزن Data Warehouse بيانات مهيكلة مُعالجة، مثل بيانات مبيعات مُنسقة، باستخدام Schema-on-Write وعملية ETL. وهو مُصمم لمحللي الأعمال لإنشاء تقارير، بتكلفة أعلى وأداء أسرع، ولكنه محدود المرونة. بينما يخزن Data Lake جميع أنواع البيانات (مهيكلة، شبه مهيكلة، وغير مهيكلة) بشكلها الخام، باستخدام Schema-on-Read و عملية ELT. وهو مُوجه لعلماء البيانات لاستكشاف البيانات وتطوير نماذج، بتكلفة أقل ومرونة عالية، لكن أدائه أقل للاستعلامات العشوائية. يُستخدم Data Warehouse لتحليل أعمال منظّم، بينما يُستخدم Data Lake في مشاريع تعلم الآلة وتحليل البيانات الضخمة. يُشكل Data Lakehouse حلاً هجيناً يجمع مزايا كليهما. اختيار الأنسب يعتمد على احتياجات المؤسسة: Data Warehouse للتقارير السريعة، و Data Lake للبيانات المتنوعة والمشاريع المعقدة.

lakhasly.com © تم تلخيص النص بواسطة موقع لخصلي