

Analiză paralelă - Balint Franceska

Data Lake vs Data Warehouse

1. Stocarea datelor

Data Lake conține toate datele unei organizații într-o formă brută, nestructurată și poate stoca datele pe termen nelimitat - pentru utilizare imediată sau viitoare.

Data Warehouse conține date structurate care au fost curățate și procesate, gata pentru analiza strategică pe baza nevoilor de afaceri predefinite.

2. Utilizatori

Datele dintr-un Data Lake - cu volumul său mare de date nestructurate - sunt utilizate de obicei de oamenii de știință de date și inginerii care preferă să studieze datele în forma lor brută pentru a obține informații noi, unice de afaceri.

Data Warehouse sunt de obicei accesate de manageri și utilizatorii finali care doresc să obțină informații din KPI-urile de afaceri, deoarece datele au fost deja structurate pentru a oferi răspunsuri la întrebări predeterminate pentru analiză.

3. Analiza

Data Lake - Analiză predictivă, învățare automată, vizualizare de date, BI, analiză de date mari.

Data Warehouse - Vizualizarea datelor, BI, analiza datelor.

4. Schema

Data Lake - Schema este definită după ce datele sunt stocate într-un lac de date vs un depozit de date, ceea ce face procesul de captare și stocare a datelor mai rapid.

Într-un Data Warehouse, schema este definită înainte ca datele să fie stocate. Acest lucru prelungește timpul necesar procesării datelor, dar odată finalizate, datele sunt pregătite pentru utilizare consecventă și sigură în întreaga organizație.

5. Prelucrare

Data Lake - ELT (Extract, Load, Transform). În acest proces, datele sunt extrase din sursa lor pentru stocare în lacul de date și structurate doar atunci când este necesar.

Data Warehouse - ETL (Extract, Transform, Load). În acest proces, datele sunt extrase din sursa (sursele), curățate, apoi structurate, astfel încât să fie gata pentru analiza la sfârșitul afacerii.

6. Cost

Costurile de stocare sunt destul de ieftine într-un Data Lake versus un Data Warehouse. Gestionarea lacurilor de date necesită, de asemenea, mai puțin timp, ceea ce reduce

OLAP vs OLTP

Arhitectură

OLAP - Utilizează data cube model care se extinde pe modelul rând cu coloană utilizat de majoritatea bazelor de date tradiționale. Acest model permite interogarea datelor multidimensionale.

OLTP - Folosește modelul convențional al bazei de date relaționale rând cu coloană.

Scop

OLAP - Colectează și extrage date agregate pentru a efectua interogări complexe pe date multidimensionale pentru a conduce analize și decizii de afaceri.

OLTP - Găsește majoritatea operațiunilor tranzacționale zilnice prin crearea, actualizarea sau ștergerea înregistrărilor, cum ar fi comenzile de coș, informațiile de livrare și multe altele.

Performanță

OLAP - Optimizat pentru sarcini de lucru grele de citire pentru efectuarea de interogări complexe pe volume mari de date. Timpul de procesare a interogărilor depinde de volumul și complexitatea interogării.

OLTP - Optimizat pentru sarcini de lucru grele de scriere și au timpi mari de procesare pentru procesarea tranzacțiilor de înaltă frecvență și volum mare, fără a compromite integritatea datelor.

Actualizări ale frecvenței datelor

OLAP - Actualizările datelor variază; unele pot apărea zilnic, săptămânal sau lunar în loturi, în funcție de nevoia de analiză.

OLTP - Actualizările au loc în timp real cu procesarea fluxului și sunt de obicei declanșate de acțiunile dvs. sau ale utilizatorilor dvs. De exemplu, eliminarea unui articol din coș ar trebui să actualizeze imediat coșul din baza de date OLTP.

Surse de date

OLAP - Mai multe baze de date OLAP acționează ca sursă de date pentru un sistem OLAP. Sursa OLAP ar putea fi, de asemenea, un depozit de date precum un depozit de date.

OLTP - Sursa de date este de obicei un sistem de baze de date relaționale.

Disponibilitate

OLAP - Disponibilitatea are prioritate scăzută pentru OLAP, deoarece este optimizată pentru analiză. De asemenea, backup-urile sunt mai puțin frecvente, deoarece se ocupă în principal de date agregate și nu modifică datele curente.

OLTP - Disponibilitatea ridicată și backup-urile frecvente sunt necesare pentru sistemele OLTP, deoarece acestea actualizează și modifică datele tranzacționale și operaționale.