

## Analiză Paralelă - Holtea Delia-Georgiana

### OLTP VS OLAP

	OLTP	OLAP
Caracteristici	Gestionează un număr mare de tranzacții mici	Gestionează volume mari de date cu interogări complexe
Tipuri query	Standard, simple	Complexe
Operații	Bazate pe INSERT, UPDATE, DELETE	Bazate pe SELECT pentru a agrega date pentru raportare
Timp de răspuns	Milisecunde	Secunde/Minute/Ore în funcție de volumul de date procesat
Design	Specific industriei, cum ar fi comerțul cu amănuntul, producția sau serviciile bancare	Specific subiectului, cum ar fi vânzările, inventarul sau marketingul
Surse	Tranzacții	Date agregate din tranzacții
Scop	Control și execută operațiuni esențiale de afaceri în timp real	Planifică, rezolvă probleme, sprijină deciziile, descoperă perspective ascunse
Data update	Actualizări scurte și rapide inițiate de utilizator	Datele sunt reîmprospătate periodic cu joburi batch programate, de lungă durată
Spațiu	În general mic, dacă datele istorice sunt arhivate	În general mare, datorită agregării seturilor mari de date
Back-up, recuperare	Sunt necesare backup-uri regulate pentru a asigura continuitatea afacerii și pentru a îndeplini cerințele legale și de guvernanță	Datele pierdute pot fi reîncărcate din baza de date OLTP, după cum este necesar, în locul copiilor de rezervă obișnuite
Productivitate	Crește productivitatea utilizatorilor finali	Crește productivitatea managerilor de afaceri, a analiștilor de date și a directorilor
Data view	Enumeră tranzacțiile de afaceri de zi cu zi	Vedere multidimensională a datelor întreprinderii
Exemple utilizatori	Personal orientat către clienți, funcționari, cumpărători online	Personal orientat către clienți, funcționari, cumpărători online
Database Design	Baze de date normalizate pentru eficiență	Baze de date normalizate pentru eficiență

## Data Lake VS Data Warehouse

	Data Lake	Data Warehouse
<b>Data Storage</b>	Conține toate datele unei organizații într-o formă brută, nestructurată și poate stoca datele pe termen nelimitat - pentru utilizare imediată sau viitoare	Conține date structurate care au fost curățate și procesate, gata pentru analiza strategică pe baza nevoilor de afaceri predefinite
<b>Utilizatori</b>	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	Datele dintr-un 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
<b>Analiza</b>	Analiză predictivă, Machine Learning, vizualizare de date, BI, analiză de big data	Vizualizarea datelor, BI, analiza datelor
<b>Schema</b>	Schema este definită după ce datele sunt stocate într-un data lake, ceea ce face procesul de captare și stocare a datelor mai rapid	Într-un depozit de date, 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
<b>Procesare</b>	ETL (Extract, Transform, Load). În acest proces, datele sunt extrase din sursa lor pentru stocare în lacul de date și structurate doar atunci când este necesar	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
<b>Costuri</b>	Costurile de stocare sunt destul de ieftine. Gestionarea lor necesită, de asemenea, mai puțin timp, ceea ce reduce costurile operaționale	Data warehouse-urile costă mai mult decât lacurile de date și necesită, de asemenea, mai mult timp pentru a fi gestionate, rezultând costuri operaționale suplimentare