

Analiză Paralelă - Tuță_Bogdan

DATA WAREHOUSE VS DATA LAKE

DEFINIRE TERMENI:

1. **Data Warehouse** - reprezintă o bază de date relațională în care sunt stocate date din sisteme tranzacționale care pot fi analizate în vederea luării celor mai bune decizii, specialiștii accesând datele prin instrumente de Business Intelligence, clienți SQL și alte aplicații de analiză.
2. **Data Lake** - este definit de asemenea ca un spațiu de stocare de date, fie că sunt structurate sau nestructurate, în cadrul căruia datele pot fi stocate așa cum sunt, fără a fi nevoie să fie structurate și rulate diferite tipuri de analize, precum de la tablouri de bord și vizualizări la procesarea BigData, analize în timp real și Machine Learning pentru a ghida la cele mai bune decizii.

PARTICULARITĂȚI DE DIFERENȚIERE:

1. Data Warehouse:

- *Tip de date stocate:* Date structurate, sistemul solicită realizarea unei scheme conceptuale înainte de a salva datele.
- *Preprocesare:* Înainte de a se realiza operațiunea de stocare trebuie realizată preprocesare cu ajutorul unor instrumente de tip ETL (Extract, Transform, Load) pentru a filtra și curăța setul de date.
- *Schema conceptuală:* Design-ul acesteia este realizat înainte de implementarea Data Warehouse, dar poate fi efectuat la momentul analizei.
- *Costuri de funcționare:* Costurile sunt mai mari decât alternativa de Data Lake.

2. Data Lake:

- *Tip de date stocate:* Date brute, date structurate, semi - structurate și nestructurate, Data Lake fiind flexibil din acest punct de vedere.
- *Preprocesare:* Preprocesarea nu este necesară, dat fiind faptul că pot fi stocate orice tipuri de date, instrumente ETL fiind utilizate ulterior dacă este necesar.
- *Schema conceptuală:* Este realizată la momentul analizei, după implementarea Data Lake.
- *Costuri de funcționare:* Este o alternativă ce pune accentul de volumul de stocare și cost în detrimentul performanței.

OLAP VS OLTP

DEFINIRE TERMENI:

1. **OLAP (Online Analytical Processing)** este fundația tehnologică a multor instrumente de business intelligence, care poate răspunde rapid la interogările de baze de date analitice multidimensionale și oferă vizualizări nelimitate ale rapoartelor, analize predictive și alte calcule analitice complexe.

2. **OLTP (Online Transactional Processing)** este un tip de procesare a datelor care execută sarcini axate pe tranzacții, cum ar fi inserarea, ștergerea sau actualizarea datelor bazei de date, fiind adesea folosit pentru tranzacții financiare, introducerea comenzilor sau vânzări cu amănuntul.

PARTICULARITĂȚI DE DIFERENȚIERE:

1. OLAP:

- *Viteză de execuție:* OLAP stochează seturi de date mai mari decât OLTP, așa că viteza este mai scăzută decât în cazul OLTP, putând varia între o secundă și câteva ore, în funcție de sistem.
- *Public țintă vizat:* Sistemele de tip OLAP sunt orientate spre afaceri și sunt utilizate de data scientists, analiști și de persoanele din conducerea unor întreprinderi, precum liderii de echipe sau directorii.
- *Frecvența de backup și update:* Sistemele OLAP procesează interogări ce pot include milioane de înregistrări de date, așadar setul de date este actualizat la oră sau zilnic, în funcție de necesitățile companiei.
- *Scop principal:* Servește scopului de a extrage informație pentru analize și procese de realizare a deciziilor organizaționale.

2. OLTP:

- *Viteză de execuție:* Timp de răspuns foarte rapid, măsurat în milisecunde.
- *Public țintă vizat:* Sistemele de tip OLTP sunt orientate către clienți, fiind concepute pentru a fi utilizat de angajați în funcții obișnuite, precum specialiștii în rezervare de hoteluri sau persoanele care cumpără online.
- *Frecvența de backup și update:* Sistemele OLTP actualizează câteva înregistrări de date simultan în timp real sau aproape în timp real, acestea realizând backup-uri mult mai frecvent decât sistemele OLAP.
- *Scop principal:* Servește misiunii de a introduce, actualiza și șterge informație dintr-o bază de date existentă.