

УДК 691.544

А. С. Луговий; Є. В. Тиш, к.т.н.

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

МЕТОДИ ОПТИМІЗАЦІЇ ПРОДУКТИВНОСТІ ETL-СИСТЕМ У БАГАТОРІВНЕВИХ АНАЛІТИЧНИХ ПЛАТФОРМАХ

UDC 691.544

A. S. Lugovyi; I. V. Tysh, Ph.D.

METHODS FOR OPTIMIZING ETL SYSTEM PERFORMANCE IN MULTI-TIER ANALYTICAL PLATFORMS

ETL (Extract, Transform, Load) є ключовим процесом для забезпечення інтеграції та аналізу великих даних у сучасних інформаційних системах. У контексті зростання обсягів даних і вимог до їх обробки, критично важливим є розробка методів оптимізації ETL-процесів, що дозволяють ефективно працювати в умовах високих навантажень та забезпечувати масштабованість системи.

Одним із ключових напрямків розвитку ETL-процесів є використання розподілених обчислювальних платформ, таких як Apache Spark або AWS Glue, що дозволяють паралельно обробляти великі обсяги даних. Це досягається шляхом поділу даних на партиції та виконання операцій одночасно на кількох вузлах кластеру.

Оптимізація також охоплює автоматизацію етапів трансформації даних. Використання метаданих для динамічного визначення схем і структури даних значно скорочує час налаштування процесів. Крім того, застосування алгоритмів для зменшення обсягів проміжних даних (наприклад, агрегування на етапі завантаження) дозволяє знизити затримки при передачі інформації.

Важливою складовою є інтеграція інструментів моніторингу та управління процесами ETL. Вони забезпечують своєчасне виявлення вузьких місць у системі та дозволяють адаптуватися до змін у робочих навантаженнях. Наприклад, адаптивне масштабування ресурсів у хмарних платформах дозволяє автоматично виділяти додаткові ресурси у пікові періоди.

З огляду на стрімке зростання обсягів даних і складність сучасних систем, однією з перспективних тенденцій є інтеграція штучного інтелекту (ШІ) для автоматизації оптимізації ETL-процесів. Використання ШІ дозволяє прогнозувати пікові навантаження, адаптувати конфігурацію процесів у реальному часі та забезпечувати виявлення аномалій. Іншою важливою тенденцією є перехід до потокових ETL-систем, які дозволяють обробляти дані в режимі реального часу, що особливо актуально для IoT і фінансових систем. Удосконалення інструментів інтеграції хмарних сервісів також відіграє важливу роль, забезпечуючи гнучкість і доступність ресурсів для ETL-процесів.

Реалізація сучасних методів оптимізації ETL-процесів у масштабованих системах обробки великих даних забезпечує підвищення продуктивності, зниження витрат на обробку даних і забезпечення їхньої доступності для бізнес-аналітики та інших потреб організації.

Література

1. Pavlo A., Aslett M. Big Data Analytics for Data Engineers. Springer, 2019. 82 с
2. Netezza, Big Data Integration: Efficient ETL Processes in a Distributed Environment. IBM Press, 2021. 382 с.