

ТЕХНОЛОГІЇ BIG DATA АНАЛІТИКИ В РОЗПОДІЛНИХ СИСТЕМАХ ОБЧИСЛЕНЬ

В сучасному світі дані постійно створюються споживачами Інтернет послуг, користувачами соціальних мереж, датчиками, реактивними двигунами, фондовими ринками та практично всім, що пов'язано з мережею. Ці дані зростають кількісно та змінюються у реальному часі. Аналіз даних також повинен здійснюватися в режимі реального часу під час збирання даних. Обсяг даних, які необхідно зберігати та аналізувати, продовжує розширюватися. Поява великих наборів даних вимагає застосування ефективних методів, технологій та інфраструктури для обробки даних та перетворення їх у діючу інформацію [1].

Основними поняттями Big Data аналізу є [2]:

- кластер – для пошуку схожих груп спостережень;
- асоціація – для пошуку спільного входження значень для різних змінних;
- регресія – для кількісної оцінки відношень між варіантами однієї чи кількох змінних.

Аналітика, проведена на великих наборах даних, може надати нові можливості і непередбачені тенденції, забезпечуючи краще представлення клієнтів і ринок у цілому. Точна аналітика клієнтів, виявлення випадків шахрайства та аналіз ризиків – все це вигоди від Big Data аналітики. Ці складні обчислення вимагають не лише великих запасів даних, але й малу затримку, високу пропускну спроможність обробки потоку.

На сьогодні основними інструментами та технологіями, які використовуються для Big Data аналітики є Python, R, Scala, Apache Spark, Hadoop, MapReduce, Cassandra, Kafka, алгоритми пошуку даних, комп'ютерне навчання, статистичні методи, NoSQL.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Олещенко Л.М. Технології Big Data / Л.М. Олещенко // Проблеми інформатизації та управління, 4(60), 2017 – С.57-64
2. David Haertzen. ETL Tools // The Analytical Puzzle: Profitable Data Warehousing, Business Intelligence and Analytics. – Technics Publications, 2012. – 346 p.