

АНАЛІТИКА ВЕЛИКИХ ДАНИХ У СЛУЖБІ СХОВИЩА НА ПЛАТФОРМІ MICROSOFT AZURE

Панфьорова І. Ю., Левченко А. С.

Харківський національний університет радіоелектроніки, Харків, Україна

Сьогодні, аналізуючи інформацію за допомогою спеціальних математичних алгоритмів і штучного інтелекту, експерти отримують корисні для бізнесу дані і інтерпретують в прогнози, підкріплені наукою [1].

Актуальність даної теми обумовлена тим, що користувачам все частіше потрібні відповіді швидше, ніж можуть забезпечити традиційні сховища даних та звичні звіти. При осмисленні потреби в хмарних технологіях варто розглянути аналітичний сервіс від Microsoft Azure Databricks. Databricks виріс із багатьох причин. Насамперед через те, що він задовольняв справжню по-требу, яку могли задовольнити системи управління реляційними базами даних (RDBMS). Databricks може працювати на стандартному устаткуванні, що спрощує його використання з наявним центром обробки даних, крім того, можливе проведення аналізу даних у хмарі [2].

Метою доповіді є виконання детального аналізу вимог до сучасних хмарних сервісів, що представлені на IT-ринку, які забезпечують оперативну аналітику різноманітних потоків цифрових даних з можливістю внесення змін в дані, що аналізуються. Аналітика потоків даних дозволяє виявити нові джерела економічної значущості, надати актуальні відомості про поведінку клієнтів і виявити тенденції ринку на ранньому етапі.

В доповіді розглянуто проблематику хмарного підходу для організації збереження даних та виявлено, що подібний потік нових даних спричиняє додаткові проблеми перед IT-фахівцями.

Наведені результати показують, що для того, щоб отримати реальну користь від великих даних, необхідні правильні інструменти для збору та організації зберігання неоднорідних типів даних з різних джерел, а також для їх легкого аналізу в контексті всіх корпоративних даних. Загальноприйнятим підходом є проєктування системи, що складається з двох частин: транзакційної та аналітичної частини інформаційної системи. Якщо ці частини роботи з даними представлені як окремі системи, то це ускладнює і вносить затримки в ETL процес, оскільки фахівцям з обробки даних потрібно переміщати, очищувати, інтегрувати дані між різними системами.

Список літератури

1. Bala R., Gill B., Smith D., Wright D., Ji K. Magic Quadrant for Cloud Infrastructure and Platform Services. 2021. Режим доступу: <https://www.gartner.com/doc/reprints?id=1-2710E4VR&ct=210802&st=sb>
2. Muniswamaiah M., Agerwala T., Tappert C. Big data in cloud computing review and opportunities. International Journal of Computer Science & Information Technology. 2019. Т. 11, №4. С. 11–43. Режим доступу: <https://arxiv.org/ftp/arxiv/papers/1912/1912.10821.pdf>