

*Зубрицький В. В., магістрант,  
Науковий керівник: Коротун О. В., канд. пед. наук,  
доцент кафедри КН  
Вакалюк Т. А., д.пед.н., проф. професор кафедри ІІЗ  
Державний університет «Житомирська політехніка», м. Житомир*

## СТВОРЕННЯ СИСТЕМИ УПРАВЛІННЯ НАВЧАННЯМ З ВИКОРИСТАННЯМ МОДЕЛІ ОБРОБКИ ВЕЛИКИХ ДАНИХ

Щоденно користувачі систем управління навчанням (СУН) формують величезну кількість даних, аналіз яких міг би суттєво вплинути на хід та розвиток навчання на всій платформі. Опрацювання великої кількості даних у ручному режимі зайняло б багато часу співробітників платформ управління навчанням, а самі звіти згенеровані у ручному режимі не дали б ніякої користі, тому що можуть бути неточними та складеними з помилками. Задля уникнення таких труднощів можна розробити спеціальний сервіс для обробки згенерованих даних СУН. Згенеровані дані можна класифікувати як великі дані – набір інформації (структурованої і неструктурованої) великих розмірів, до якої традиційні способи та підходи (здебільшого засновані на рішеннях класу бізнес аналітики та системи управління базами даних) не можуть бути застосовані [2].

Для обробки даних пропонується застосувати програмну модель та програмний каркас «MapReduce». Дана програмна модель, дозволяє виконувати розподілену та паралельну обробку великих наборів даних у розподіленому середовищі, як видно з назви програмна модель складається з двох окремих завдань – Map та Reduce [1]. Основні функції програмної моделі: map – приймає на вхід список значень та деяку функцію, що потім застосовує до кожного елемента списку та повертає новий список; reduce (згортка) – перетворює список до єдиного атомарного значення за допомогою заданої функції, якою кожної ітерації передаються новий елемент списку і проміжний результат.

Обробка даних проходить в три етапи [1].

1. Map: попередня обробка вхідних даних як великого переліку значень, кожен робочий вузол застосовує функцію Map до локальних даних та записує результат у форматі «ключ-значення» у тимчасове сховище.

2. Shuffle: робочі вузли перерозподіляють дані на основі ключів, раніше створених функцією Map, таким чином, щоб усі дані одного ключа належали одному робочому вузлу.

3. Reduce: паралельна обробка кожним робочим вузлом кожної групи даних за порядком ключів. Головний вузол отримує проміжні відповіді від робочих вузлів і передає їх у вільні вузли до виконання наступного кроку. Результат, що вийшов після проходження всіх необхідних кроків – це і є рішення вихідного завдання.

Основна перевага MapReduce - легкість масштабування обробки даних на декількох обчислювальних вузлах. Декомпозиція програми обробки даних на map та reduce іноді нетривіальна. Якщо програму написати у формі MapReduce, то масштабування програми для запуску на сотнях, тисячах або навіть десятках тисяч машин у кластері – це лише зміна конфігурації. Ця проста масштабованість приваблює багатьох програмістів використовувати модель MapReduce [1]. Google представив алгоритм MapReduce для виконання масової паралельної обробки дуже великих наборів даних за допомогою кластерів серверного обладнання. Як зазначалося раніше за допомогою MapReduce можна значно спростити та скоротити час опрацювання даних отриманих під час роботи СУН, якщо в минулому для таких обчислень потрібен був суперкомп'ютер та науковці, то тепер з цим може легко впоратись відділ розробки програмного забезпечення [3]. Візьмемо як приклад, адміністратор курсу певної компанії змінив наповнення та додав нові завдання, щоб підняти загальний рівень підготовки працівників, що будуть проходити цей тест, і захотів перевірити результат навчання до і після цих змін, це ідеальна проблема в яку вписується MapReduce. У цьому випадку необхідно запустити модель регресії щодо групи працівників, які пройшли модифікований курс, і підрахувати наскільки він став ефективний.

У підсумку, хотілось би зазначити, що з появою інтернету та комунікаційних технологій поширення електронного навчання зросло і користується великим попитом. Цифрові дані, які створюються закладами вищої освіти, також на підйомі. Необхідність використання платформ «великих даних» для обробки та аналізу цих великих обсягів даних є надзвичайно важливою та актуальною. Багато навчальних систем вже використовують аналітику для покращення свого процесу.

### Список використаних джерел та літератури

1 Tutorialspoint.com. 2021. Hadoop - MapReduce. [online] Available at: <[https://www.tutorialspoint.com/hadoop/hadoop\\_mapreduce.htm](https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm)>.

2. Venkatram, K. and Geetha, M., 2017. Review on Data & Analytics – Concepts, Philosophy, Process and Applications

3. Medium. 2021. Fundamentals of MapReduce with MapReduce Example. [online] Available at: <<https://medium.com/edureka/mapreduce-tutorial-3d9535ddbe7c>> [Accessed 14 November 2021].