

РЕФЕРАТ

Магістерська дисертація: 65 сторінок, 19 рисунків, 7 таблиць, 1 додаток, 15 джерел.

Актуальність. Глобальна комп'ютеризація, прогрес в сфері мікроелектроніки та інформаційних технологій обумовлюють збільшення об'ємів інформації. Загальний об'єм даних на 2012 рік складав більше 1,8 зеттабайт (1,8 трлн Гб) та за дослідженнями IDC (International Data Corporation) ця цифра подвоюється кожні 2 роки[9]. Дев'яносто відсотків даних у світі сьогодні створено лише за останні два роки. Наш поточний вихід даних складає приблизно 2,5 мільярда Гб на день. Збереження такого об'єму інформації потребує немалих ресурсів та зусиль. Але найбільша проблема полягає не в збереженні даних, а в їх обробці - за прогнозами IDC в 2020 році з усіх об'ємів даних лише 35% будуть корисними. Таким чином потрібно механізм для спрощення даних для подальшого їх аналізу. Одним з таких методів є кластеризація даних.

Кластеризація застосовується для того, щоб зробити так зване стиснення даних, тобто скоротити обсяг використовуваних даних за рахунок того, що всередині кластера об'єкти не розрізняються (розглядаються як один об'єкт).

Тому задача автоматичної та якісної кластеризації даних за прийнятний час, без попереднього структурування та дослідження інформації, є задачею, яка варта уваги та досліджень.

Робота присвячена розробці модифікації для існуючих алгоритмів кластерного аналізу шляхом застосування паралельних розрахунків на графічному процесорі.

Мета і завдання дослідження. Метою є удосконалення існуючих методів кластеризації даних на основі технології паралельних обчислень Nvidia Cuda для прискорення їх роботи. Для досягнення мети необхідно вирішити наступні задачі:

– дослідити предметну область сучасних методів для роботи з надвеликими масивами інформації і детально розглянути методи кластеризації даних;

- аналіз технології паралельних обчислень Nvidia CUDA на графічних процесорах;
- розробити удосконалені методи кластеризації на основі паралельних обчислень на графічному процесорі за допомогою технології Nvidia Cuda;
- розробити програмний застосунок для кластеризації даних з реалізацією обраних методів та їх удосконалених варіантів;
- виконати експериментальне дослідження розробленого алгоритму.

Об'єктом дослідження є процеси кластеризації даних.

Предметом дослідження є швидкодія та ефективність методів кластеризації даних.

Зв'язок роботи з науковими програмами, планами і темами. Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924.

Наукова новизна отриманих результатів полягає в розробці та реалізації модифікованих методів кластеризації за допомогою паралельних обчислень на графічному процесорі Nvidia.

Публікації. Матеріали роботи опубліковані в тезах конференцій «ІНФОРМАТИКА ТА ОБЧИСЛЮВАЛЬНА ТЕХНІКА – ІОТ-2018» та «ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ПРИЙНЯТТЯ РІШЕНЬ ТА ПРОБЛЕМИ ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ».

КЛАСТЕРНИЙ АНАЛІЗ, ПАРАЛЕЛЬНІ ОБЧИСЛЕННЯ, ГРАФІЧНИЙ ПРОЦЕСОР, NVIDIA CUDA