

УДК 004

Муравський І.А.¹, Поляков М.О.²

¹ студ. гр. КНТ-214м НУ «Запорізька політехніка»

² проф. НУ «Запорізька політехніка»

ДОСЛІДЖЕННЯ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ ЗАСОБІВ ОБРОБКИ ВЕЛИКИХ НАБОРІВ ДАНИХ

Об'єкт дослідження – процес обробки великих наборів даних.

Предмет дослідження – методи та засоби для обробки великих наборів даних.

Метою роботи є прискорення процесу обробки великих наборів даних, шляхом паралелізації.

Матеріали, методи та технічні засоби: мова програмування Python, середовище розробки PyCharm, бібліотеки та фреймворки PySpark та TensorFlow.

В ході роботи було досліджено та розроблено нові засоби обробки великих наборів даних.

Головою метою роботи було прискорення обробки великих наборів даних. Для досягнення цієї мети було використано паралелізацію, засобами фреймворку PySpark, процесів обробки великих наборів даних.

У роботі розроблявся застосунок для обробки великих наборів даних.

Представлено дослідницьку частину роботи з результатами огляду найбільш відомих технологій та засобів обробки великих наборів даних. Під час аналізу існуючих рішень було прийнято рішення, щодо розробки нового застосунку, який би вирішував низку проблем існуючих конкурентів.

Для тестування запропонованого рішення використовувалася реальна велика вибірка даних під час опрацювання якої запропонований підхід продемонстрував прийнятні результати, зокрема і прискорення при паралельному виконанні.

Розроблено програмне забезпечення для обробки великих наборів даних на основі мови програмування Python, середовища розробки PyCharm, бібліотеки PySpark та фреймворка TensorFlow.

На основі досліджених переваг та недоліків засобів обробки великих наборів даних та виконаного проектування нового рішення, розроблено новий засіб для обробки великих наборів даних. Здійснено тестування розробленого програмного забезпечення для обробки великих наборів даних.

За результатами роботи слід зробити висновок про те, що нове рішення обробки великих наборів даних продемонструвало хороші результати.

В якості можливих напрямків розвитку можна розглянути такі варіанти покращення результатів:

- тестування розробленого застосунку на кластерній обчислювальній системі, замість просто багатоядерної обчислювальної системи;
- використання в подальшому мови Java та механізму Java Virtual Machine для оптимізації та пришвидшення роботи застосунку в цілому.

Новизна роботи полягає в тому, що розроблено новий підхід до обробки великих наборів даних.

Практичне значення роботи полягає в тому, що за рахунок використання паралелізації, вдалося пришвидшити обробку великих наборів даних.

Галузь використання – обробка великих та надмірних даних, онлайн-комерція, промислові виробництва.